Revised proposal to encode Hanifi Rohingya in Unicode

Anshuman Pandey
pandey@umich.edu

December 31, 2016

1 Introduction

This is a proposal to encode the Hanifi Rohingya script in Unicode. It supersedes the following documents:

- L2/12-214 “Preliminary Proposal to Encode the Rohingya Script”
- L2/12-278R “Proposal to encode the Hanifi Rohingya script in Unicode”

A proposal summary form is enclosed. The present document addresses comments made in:

- L2/16-037 “Recommendations to UTC #146 January 2016 on Script Proposals”
- L2/17-037 “Recommendations to UTC #150 January 2017 on Script Proposals”

2 Background

The ‘Hanifi Rohingya’ script is used for writing Rohingya (ISO 639-3: rhg), an Indo-Aryan language spoken by one million people in Myanmar (Rakhine State) and by 200,000 people in Bangladesh (Cox’s Bazaar District). The Rohingya (𐴌𐴟𐴖𐴥𐴝𐴚𐴒𐴙𐴝 *ruwainggya*, 𐴌𐴟𐴇𐴥𐴝𐴚𐴒𐴙𐴝 *ruhainggya*) language is also spoken in other countries along the Indian Ocean because of the dispersion of the Rohingya community on account of persecution in Myanmar. In addition to the present script, three others are used for writing Rohingya: Burmese, Arabic, and Latin. The Arabic orthography is known as ‘Rohingya Fonna’. The Latin adaptation, known as ‘Rohingyalish’, is based upon the English alphabet.

The script was developed by the Rohingya Language Committee in the 1980s under the guidance of Maulana Mohammed Hanif. The initial script approved by the committee is shown in the document in figure 2. The script is a modern construction and has no direct genetic affiliation to other writing systems. However, the influence of Arabic is evident in the general appearance and structure of the script, and some letter shapes are influenced by characters from the Latin and Myanmar scripts. The writing system is referred to by its users as 𐴌𐴟𐴇𐴥𐴝𐴚𐴒𐴙𐴝 𐴎𐴟𐴁𐴝𐴕𐴡𐴌 𐴇𐴟𐴌𐴟𐴉 *ruhainggya zubanor huruf* “letters of the Rohingya language”, or simply *ruhainggya huruf* “Rohingya letters”. The designation “Hanifi Rohingya” has been selected for the script block in Unicode (see § 3.2).

There is limited information available on the Hanifi Rohingya script in English. Most of the materials on the script are written in the Rohingya language. The Rohingya Language Committee and the Rohingya Education Board Myanmar have published multilingual primers of the script (see figures 5–20). The script is used for the publication of books and newspapers, which are handwritten and printed. These sources show that a few Rohingya fonts have been developed for desktop publishing. The latest typefaces have been developed by Muhammad Noor.

3 Script Details

3.1 Structure

Hanifi Rohingya is an alphabetic script that is written from right to left. Consonant letters represent pure consonants. Vowels are expressed using letter-like signs, which are placed after a consonant. Vowel signs represent short vowels. Vowel length and stress is indicated using tonal signs, which are placed above vowel signs. An independent or word-initial vowel is represented using a vowel carrier and a vowel sign. Nasalization is indicated using a letter-like sign that is placed after a vowel sign. Consonant gemination is marked by a combining sign placed above a letter. Although letters are purely consonantal, a letter-like sign is often used optionally for indicating a bare consonant at the end of a word.

The script is structurally conjoining and is modelled upon Arabic. Letters join to adjacent letters at the baseline. In several hand-written and printed sources, the baseline between adjacent letters may not be completely connected. This is unintentional and is a result of casual hand writing or a limitation of font design. In texts printed using modern digitized fonts, the connections between letters are more consistently maintained.

Spacing is used for separating words. Latin and Arabic punctuation signs are commonly used for delimiting text segments, eg. periods, commas, colons, hyphens. The available sources do not show any specific linebreaking features, such as hyphens. An elongation feature similar to Arabic *kashida* is often used for justification.

3.2 Script name

The name 'Hanifi Rohingya' is assigned to the Unicode script block. This name has been selected because it precisely distinguishes the script from others that are used for writing Rohingya. Given that orthographies for Rohingya have been developed using other scripts, referring to the present script block using the bare name 'Rohingya' would present ambiguity.

3.3 Character names

Names for characters are based upon normalized Latin transliterations of Rohingya forms given in various script primers (see tables 1 and 2 for comparisons of names). In general, the proposed Unicode character names align with Latin names given for the majority of letters across the available Rohingya sources. In cases where there are multiple names for a character, an appropriate name has been chosen in consultation with the user community. Notes on the names for specific characters are given throughout the proposal.

For sake of brevity, the descriptor ‘HANIFI ROHINGYA’ is dropped when refering to Hanifi Rohingya characters, eg. HANIFI ROHINGYA LETTER A is referred to as A. Characters of other scripts are designated by their full Unicode names.

3.4 Transliteration

The transliteration of Hanifi Rohingya used in this document is based upon the Rohingyalish, unless specified otherwise.

3.5 Representative glyphs

The representative glyphs are derived from the “Rohingya Gonya Leyka Noories” typeface developed by Muhammad Noor. The glyphs are used with permission. Original glyphs have been modified and additional glyphs have been introduced by the proposal author.

4 Character Repertoire

The proposed repertoire for ‘Hanifi Rohingya’ contains 50 characters. There appears to be a one-to-one relationship between phonemes of the Rohinngya language and the basic letters and vowel signs of the Hanifi Rohingya script (see figure 1). The encoded set may differ from traditional and scholarly inventories of script varieties that occur in written and inscriptional sources. Such differences naturally arise from the requirements for digitally representing a script in plain text and for preserving the semantics of characters.

4.1 Basic letters

There are 28 letters, as shown below with positional forms:

	Character name	IPA	Join	Fin.	Med.	Init.
𐴀	HANIFI ROHINGYA LETTER A	/Ø/	left	—	—	𐴀
𐴁	HANIFI ROHINGYA LETTER BA	/b/	dual	𐴁	𐴁	𐴁
𐴂	HANIFI ROHINGYA LETTER PA	/p/	dual	𐴂	𐴂	𐴂
𐴃	HANIFI ROHINGYA LETTER TA	/t/	dual	𐴃	𐴃	𐴃
𐴄	HANIFI ROHINGYA LETTER TTA	/ʈ/	dual	𐴄	𐴄	𐴄
𐴅	HANIFI ROHINGYA LETTER JA	/ɟ/	dual	𐴅	𐴅	𐴅
𐴆	HANIFI ROHINGYA LETTER CA	/c/	dual	𐴆	𐴆	𐴆
𐴇	HANIFI ROHINGYA LETTER HA	/h/	dual	𐴇	𐴇	𐴇
𐴈	HANIFI ROHINGYA LETTER KHA	/x/	dual	𐴈	𐴈	𐴈

𐴉	HANIFI ROHINGYA LETTER FA	/f/	dual	𐴉‍	‍𐴉‍	‍𐴉
𐴊	HANIFI ROHINGYA LETTER DA	/d/	dual	𐴊‍	‍𐴊‍	‍𐴊
𐴋	HANIFI ROHINGYA LETTER DDA	/ɖ/	dual	𐴋‍	‍𐴋‍	‍𐴋
𐴌	HANIFI ROHINGYA LETTER RA	/r/	dual	𐴌‍	‍𐴌‍	‍𐴌
𐴍	HANIFI ROHINGYA LETTER RRA	/ɽ/	dual	𐴍‍	‍𐴍‍	‍𐴍
𐴎	HANIFI ROHINGYA LETTER ZA	/z/	dual	𐴎‍	‍𐴎‍	‍𐴎
𐴏	HANIFI ROHINGYA LETTER SA	/s/	dual	𐴏‍	‍𐴏‍	‍𐴏
𐴐	HANIFI ROHINGYA LETTER SHA	/ʃ/	dual	𐴐‍	‍𐴐‍	‍𐴐
𐴑	HANIFI ROHINGYA LETTER KA	/k/	dual	𐴑‍	‍𐴑‍	‍𐴑
𐴒	HANIFI ROHINGYA LETTER GA	/g/	dual	𐴒‍	‍𐴒‍	‍𐴒
𐴓	HANIFI ROHINGYA LETTER LA	/l/	dual	𐴓‍	‍𐴓‍	‍𐴓
𐴔	HANIFI ROHINGYA LETTER MA	/m/	dual	𐴔‍	‍𐴔‍	‍𐴔
𐴕	HANIFI ROHINGYA LETTER NA	/n/	dual	𐴕‍	‍𐴕‍	‍𐴕
𐴖	HANIFI ROHINGYA LETTER WA	/w/, /v/	dual	𐴖‍	‍𐴖‍	‍𐴖
𐴗	HANIFI ROHINGYA LETTER KINNA WA	/u/	dual	𐴗‍	‍𐴗‍	‍𐴗
𐴘	HANIFI ROHINGYA LETTER YA	/j/	dual	𐴘‍	‍𐴘‍	‍𐴘
𐴙	HANIFI ROHINGYA LETTER KINNA YA	/i/	dual	𐴙‍	‍𐴙‍	‍𐴙
𐴚	HANIFI ROHINGYA LETTER NGA	/ŋ/	dual	𐴚‍	‍𐴚‍	‍𐴚
𐴛	HANIFI ROHINGYA LETTER NYA	/ɲ/	dual	𐴛‍	‍𐴛‍	‍𐴛

Only the representative form of a letter is encoded (left column in the above). This form is identical to the isolated or independent shape of a letter. The positional forms will exist as glyphs in a font and will be substituted for the isolated letters by the shaping engine (see § 5.1).

Although consonant letters represent pure consonants, the character names follow the Unicode naming conventions for South Asian scripts derived from Brahmi. For instance, 𐴁 is 'BA' not 'B'. This convention aligns with the transliterations for letters given in script charts and with the pronunciations of letter names in the Rohingya language. The letters 𐴚 NGA and 𐴛 NYA are also known as 𐴒𐴝𐴕 *gan* and 𐴕𐴝𐴘𐴧 *nayya*, respectively, and these names are specified as aliases in the names list.

4.1.1 Notes on letters

𐴀 A In script charts, the letter 𐴀 is given Latin names such as *a*, *aa*. However, it does not represent /ɔ/ or /a/ nor does it have an independent value. It is a consonant placeholder modeled upon ا U+0627 ARABIC

LETTER ALEF and functions as a carrier for vowel signs for representing independent or word-initial vowels. Although it may be appropriate to assign the letter a name such as 'vowel carrier', it has been given the name 'A' because of user preference.

𐴊 DA The 𐴊 is used in early materials for isolated and word-final forms of 𐴊 DA (see figures 5, 17). It inherently represents a bare consonant and cannot be followed by SAKIN. It is used in script charts (see figures 2, 5, 7) and in running text. However, as described in figure 3, 𐴊 was replaced by 𐴊𐴢 <𐴊 DA, 𐴢 SAKIN>, as seen in the charts in figures 6, 8, 9. This 𐴊 may be considered a glyphic variant of the ligature 𐴊𐴢 and may be represented by substituting the sequence <𐴊 DA, 𐴢 SAKIN> with a dedicated glyph in a font.

𐴋 DDA The 𐴋 is an older form of 𐴋 DDA (see explanation in figures 3 and 4). It was revised into the present form in order to prevent confusion with other letters, ie. 𐴍 RRA.

𐴓 LA The letter 𐴓 LA has a stylized version 𐴓 in which the terminal is curled (see figure 5). When used, this form of LA appears only in isolated and final positions (see figure 16).

𐴔 MA This is the only letter whose shape changes significantly when it occurs in initial and medial positions. The vertical descender of 𐴔 MA is retained in the final and isolated forms.

𐴗 KINNA WA The letter 𐴗 𐴑𐴞𐴕𐴕𐴝 *kinna wa*, or "small *wa*", is used for representing diphthongs (see § 5.4).

𐴙 KINNA YA The letter 𐴙 𐴑𐴞𐴕𐴕𐴝 *kinna ya*, or "small *ya*", is used for representing diphthongs (see § 5.4).

4.2 Additional letter

The following letter is also included in the repertoire, as shown below with positional forms:

	Character name	IPA	Join	Fin.	Med.	Init.
𐴜	HANIFI ROHINGYA LETTER VA	/v/	dual	𐴜	𐴜	𐴜

The letter 𐴜 was created by some users for representing /v/ in the transliteration of loanwords, eg. 𐴜𐴠𐴕 *ven* for English "van" (see figure 18). It was not used widely and its usage has been discontinued. The letter 𐴖 WA is now used for indicating both /w/ and /v/. Nonetheless, the letter VA is encoded in order to represent existing documents that contain the letter. This letter has been added to the repertoire with approval of the user community.

4.3 Vowel signs

The following five vowel signs are used for representing basic vowel sounds (see § 5.3):

	Character name	Value	Join	Final	Medial	Initial
𐴝	HANIFI ROHINGYA VOWEL SIGN A	/ɔ/, /a/	dual	𐴝	𐴝	𐴝
𐴞	HANIFI ROHINGYA VOWEL SIGN I	/i/	dual	𐴞	𐴞	𐴞

𐴟	HANIFI ROHINGYA VOWEL SIGN U	/u/	dual	‍𐴟	‍𐴟‍	𐴟‍
𐴠	HANIFI ROHINGYA VOWEL SIGN E	/e/	dual	‍𐴠	‍𐴠‍	𐴠‍
𐴡	HANIFI ROHINGYA VOWEL SIGN O	/o/	dual	‍𐴡	‍𐴡‍	𐴡‍

Indigenous names for these signs are 𐴝𐴉𐴡𐴌 *a-for*, 𐴞𐴉𐴡𐴌 *i-for*, 𐴟𐴉𐴡𐴌 *u-for*, 𐴠𐴉𐴡𐴌 *e-for*, 𐴡𐴉𐴡𐴌 *o-for*, in which 𐴉𐴡𐴌 *for* means “sign”. These names are specified as aliases in the names list. The script does not have dedicated signs for long vowels. Instead, length is indicated using tonal signs (see § 4.6). In valid orthography, vowel signs can only occur after a letter, and never in initial position. However, the isolated forms appear in script primers for purposes of illustration.

4.4 Vowel silencer

	Character name	Value	Join	Final	Medial	Initial
𐴢	HANIFI ROHINGYA SAKIN	—	right	‍𐴢	—	—

The sign SAKIN is used for indicating a bare consonant, usually when a letter is shown in isolation or after a letter at the end of a word (see § 5.2, and figure 16 for examples). The character name is based upon the Hanifi Rohingya spelling of the indigenous name: 𐴏𐴝𐴑𐴞𐴕 *sakin*, also 𐴏𐴝𐴦𐴑𐴞𐴕 *sáakin*. The sign SAKIN is similar in function to ◌ْ U+0652 ARABIC SUKUN. The SAKIN is a right-joining letter, and is only used after a letter in valid orthography. It is used in isolated contexts in script primers for purposes of illustration.

4.5 Nasalization mark

	Character name	Value	Join	Final	Medial	Initial
𐴣	HANIFI ROHINGYA NA KHONNA	/n/	dual	‍𐴣	‍𐴣‍	𐴣‍

The sign NA KHONNA is used for indicating nasalization (see § 5.5). The character name is based upon the Hanifi Rohingya spelling of the indigenous name: 𐴕𐴝𐴈𐴡𐴕𐴧𐴝 *na khonna*. In valid orthography, the NA KHONNA can only be used after a vowel sign or a tonal sign. Although it could theoretically be used in word-initial position, such usage does not reflect actual phenomena.

4.6 Tonal signs

	Character	tone type
◌𐴤	HANIFI ROHINGYA SIGN HARBAY	short high tone
◌𐴥	HANIFI ROHINGYA SIGN TELA	long rising tone
◌𐴦	HANIFI ROHINGYA SIGN TANA	long falling tone

These combining signs are used for representing tones (see § 5.3): The character names are based upon the Hanifi Rohingya spellings of the indigenous names: 𐴇𐴝𐴤𐴌 𐴁𐴝𐴤𐴘 *háar báy*; 𐴃𐴠𐴓𐴝 *tela*; 𐴃𐴝𐴕𐴝 *tana*. In valid orthography these tonal signs can only combine with a vowel sign.

4.7 Gemination sign

	Character
◌𐴧	HANIFI ROHINGYA SIGN TOSSI

The ◌𐴧 SIGN TOSSI is used for indicating geminated (doubled) consonants (see § 5.6). The character name is based upon the Hanifi Rohingya spelling of the indigenous name: 𐴃𐴡𐴏𐴧𐴞 *tossi*. The sign is identical in function to ◌ّ U+0651 ARABIC SHADDA. In valid orthography, the sign can only combine with a letter.

4.8 Digits

There is a full set of decimal digits:

	Character	Rohingya name		Value
𐴰	HANIFI ROHINGYA DIGIT ZERO	𐴐𐴡𐴕𐴧𐴡	*shonno*	0
𐴱	HANIFI ROHINGYA DIGIT ONE	𐴊𐴟𐴘	*duy*	1
𐴲	HANIFI ROHINGYA DIGIT TWO	𐴃𐴞𐴕	*tin*	2
𐴳	HANIFI ROHINGYA DIGIT THREE	𐴆𐴝𐴘𐴌	*chayr*	3
𐴴	HANIFI ROHINGYA DIGIT FOUR	𐴉𐴝𐴕𐴏	*fans*	4
𐴵	HANIFI ROHINGYA DIGIT FIVE	𐴀𐴠𐴑	*ek*	5
𐴶	HANIFI ROHINGYA DIGIT SIX	𐴏𐴡𐴤	*só*	6
𐴷	HANIFI ROHINGYA DIGIT SEVEN	𐴇𐴝𐴤𐴕𐴃	*hánt*	7
𐴸	HANIFI ROHINGYA DIGIT EIGHT	𐴀𐴝𐴕𐴐𐴃	*ansht*	8
𐴹	HANIFI ROHINGYA DIGIT NINE	𐴕𐴡	*no*	9

5 Script Details

5.1 Shaping model

The shaping model for Rohingya is based upon that for Arabic. Each character is placed in the logical order in the input stream. For letters, the isolated form is substituted by the appropriate positional form. In the

examples below the 'input' column shows the input stream. The 'shaped glyphs' show the substitutions for each letter. The 'output' depicts the expected display.

Input →	Shaped glyphs ←	Output ←
𐴁 BA, 𐴁 BA, 𐴁 BA	𐴁 + 𐴁 + 𐴁	𐴁𐴁𐴁
𐴑 KA, 𐴃 TA, 𐴁 BA	𐴑 + 𐴃 + 𐴁	𐴑𐴃𐴁
𐴇 HA, 𐴌 RA, 𐴉 FA	𐴇 + 𐴌 + 𐴉	𐴇𐴌𐴉

In general, positional forms of letters are identical to their isolated shape. The difference is typographic: positional glyphs have edges at the baseline for facilitating connections.

The letter 𐴔 MA differs from other letters because its isolated form is distinct from its initial form:

Input →	Shaped glyphs ←	Output ←
𐴔 MA, 𐴔 MA, 𐴔 MA	𐴔 + 𐴔 + 𐴔	𐴔𐴔𐴔
𐴏 SA, 𐴓 LA, 𐴔 MA	𐴏 + 𐴓 + 𐴔	𐴏𐴓𐴔

There are sources in which the final form of 𐴔 MA is rendered similar to the initial shape. Producing this form requires breaking regular shaping behavior using the generic control character ‍ U+200D ZERO WIDTH JOINER (abbreviated ZWJ). The ZWJ may be placed after MA to produce a final form without the ascender:

Input →	Shaped glyphs ←	Output ←
𐴔 MA, 𐴔 MA	𐴔 + 𐴔	𐴔𐴔
𐴔 MA, 𐴔 MA, ‍ ZWJ	𐴔 + 𐴔	𐴔𐴔‍

Only 𐴔 MA has such distinctions in its positional forms. The ZWJ should only affect this letter and it should be ignored if placed after other letters.

5.2 Usage of the sign SAKIN

Although consonant letters represent pure consonants, the script has the sign 𐴢 SAKIN for indicating the absence of a vowel sign. The sign occurs frequently in script charts for representing consonants in isolated contexts. It is also used commonly in running text at the end of a word (see figure 16).

b	𐴁𐴢	<𐴁 BA, 𐴢 SAKIN>
p	𐴂𐴢	<𐴂 PA, 𐴢 SAKIN>
t	𐴃𐴢	<𐴃 TA, 𐴢 SAKIN>

The SAKIN is placed after a letter in the input stream. The occurrence of SAKIN in word-initial position should be considered orthographically invalid.

Usage of SAKIN is optional, and highly dependently upon user preference. There are no formal conventions regarding its use. Even in script charts the usage of SAKIN is inconsistent. For example, the chart in figure 8 shows SAKIN written with every letter (except for MA), including 𐴀 A. On the other hand, figures 5, 6, 7 show SAKIN used with only certain letters, which are enumerated below:

𐴁𐴢	𐴁	BA	𐴈𐴢	𐴈	KHA	𐴒𐴢	𐴒	GA
𐴂𐴢	𐴂	PA	𐴉𐴢	𐴉	FA	𐴓𐴢	𐴓	LA
𐴃𐴢	𐴃	TA	𐴊𐴢	𐴊	DA	𐴖𐴢	𐴖	WA
𐴄𐴢	𐴄	TTA	𐴏𐴢	𐴏	SA	𐴗𐴢	𐴗	KINNA WA
𐴇𐴢	𐴇	HA	𐴐𐴢	𐴐	SHA	𐴙𐴢	𐴙	KINNA YA

The SAKIN was introduced by some users after the script was developed. Indeed, the chart showing the finalized script as approved by the committee led by Maulana Mohammad Hanif in 1980 does not show SAKIN with any letter (see figure 2), except possibly with 𐴙 KINNA YA (𐴙𐴢).

5.3 Representation of vowels

Independent and word-initial vowels are represented using the letter 𐴀 A followed by a vowel sign:

	Display	Input →
a	𐴀𐴝	𐴀 A, 𐴝 VOWEL SIGN A
i	𐴀𐴞	𐴀 A, 𐴞 VOWEL SIGN I
u	𐴀𐴟	𐴀 A, 𐴟 VOWEL SIGN U
e	𐴀𐴠	𐴀 A, 𐴠 VOWEL SIGN E
o	𐴀𐴡	𐴀 A, 𐴡 VOWEL SIGN O

Non-initial vowel signs are represented with consonants in the same manner as with 𐴀 A:

	Display	Input →
b	𐴁	𐴁 BA
ba	𐴁𐴝	𐴁 BA, 𐴝 VOWEL SIGN A
bi	𐴁𐴞	𐴁 BA, 𐴞 VOWEL SIGN I
bu	𐴁𐴟	𐴁 BA, 𐴟 VOWEL SIGN U
be	𐴁𐴠	𐴁 BA, 𐴠 VOWEL SIGN E
bo	𐴁𐴡	𐴁 BA, 𐴡 VOWEL SIGN O

Vowel length and stress are indicated using tonal signs. A tonal sign is placed after a vowel sign in the input stream. A short high tone is indicated using ◌𐴤 SIGN HARBAY as follows:

	Display	Input →
bá	𐴁𐴝𐴤	𐴁 BA, 𐴝 VOWEL SIGN A, ◌𐴤 SIGN HARBAY
bí	𐴁𐴞𐴤	𐴁 BA, 𐴞 VOWEL SIGN I, ◌𐴤 SIGN HARBAY
bú	𐴁𐴟𐴤	𐴁 BA, 𐴟 VOWEL SIGN U, ◌𐴤 SIGN HARBAY
bé	𐴁𐴠𐴤	𐴁 BA, 𐴠 VOWEL SIGN E, ◌𐴤 SIGN HARBAY
bó	𐴁𐴡𐴤	𐴁 BA, 𐴡 VOWEL SIGN O, ◌𐴤 SIGN HARBAY

A long rising tone is expressed using the ◌𐴥 SIGN TELA, as follows:

	Display	Input →
baá	𐴁𐴝𐴥	𐴁 BA, 𐴝 VOWEL SIGN A, ◌𐴥 SIGN TELA
bií	𐴁𐴞𐴥	𐴁 BA, 𐴞 VOWEL SIGN I, ◌𐴥 SIGN TELA
buú	𐴁𐴟𐴥	𐴁 BA, 𐴟 VOWEL SIGN U, ◌𐴥 SIGN TELA
beé	𐴁𐴠𐴥	𐴁 BA, 𐴠 VOWEL SIGN E, ◌𐴥 SIGN TELA
boó	𐴁𐴡𐴥	𐴁 BA, 𐴡 VOWEL SIGN O, ◌𐴥 SIGN TELA

A long falling tone is indicated using the ◌𐴦 SIGN TANA, as follows:

	Display	Input →
báa	𐴁𐴝𐴦	𐴁 BA, 𐴝 VOWEL SIGN A, ◌𐴦 SIGN TANA
bíi	𐴁𐴞𐴦	𐴁 BA, 𐴞 VOWEL SIGN I, ◌𐴦 SIGN TANA
búu	𐴁𐴟𐴦	𐴁 BA, 𐴟 VOWEL SIGN U, ◌𐴦 SIGN TANA
bée	𐴁𐴠𐴦	𐴁 BA, 𐴠 VOWEL SIGN E, ◌𐴦 SIGN TANA
bóo	𐴁𐴡𐴦	𐴁 BA, 𐴡 VOWEL SIGN O, ◌𐴦 SIGN TANA

As is evident in the available sources, the placement of tonal signs may vary (see figure 15). The normative position is between the consonant and vowel sign. But, in some sources they appear above the vowel sign. Variation of placement may be considered a stylistic matter. Regardless of the position in the rendered output, in the in the input stream the tonal sign is to be placed after the vowel sign. The font should manage the proper positioning of the tonal signs.

Other stylistic issues regarding tonal signs include the size of the sign. The ◌𐴦 SIGN TANA may be elongated to span the letter and vowel sign, as shown below. This is also to be managed in the font.

	Display	Input →
báa	𐴁𐴝𐴦	𐴁 BA, 𐴝 VOWEL SIGN A, ◌𐴦 SIGN TANA
	𐴁𐴝𐴦	𐴁 BA, 𐴝 VOWEL SIGN A, ◌𐴦 SIGN TANA

Although script charts show vowel signs in isolation without the vowel-carrier letter 𐴀 A, these cases simply illustrate the character and are not depictions of valid usage (see, for example, figure 5). In valid orthography each vowel sign must be preceded by a consonant base. Multiple vowel signs on a single base are not considered valid sequences.

		Display	Input →
bae	correct	𐴁𐴝𐴀𐴠	𐴁 BA, 𐴝 VOWEL SIGN A, 𐴀 A, 𐴠 VOWEL SIGN E
	incorrect	𐴁𐴝𐴠	𐴁 BA, 𐴝 VOWEL SIGN A, 𐴠 VOWEL SIGN E
bao	correct	𐴁𐴝𐴀𐴡	𐴁 BA, 𐴝 VOWEL SIGN A, 𐴀 A, 𐴡 VOWEL SIGN O
	incorrect	𐴁𐴝𐴡	𐴁 BA, 𐴝 VOWEL SIGN A, 𐴡 VOWEL SIGN O

5.4 Representation of diphthongs

The letters 𐴙 KINNA YA and 𐴗 KINNA WA are used for representing diphthongs. The KINNA YA 'small *ya*' represents a semi-vowel glide (/j/, /i/) and KINNA WA 'small *wa*' represents the glide (/w/, /u/). They are placed after vowel signs to indicate a diphthong.

	Display	Input →
bai	𐴁𐴝𐴙	𐴁 BA, 𐴝 VOWEL SIGN A, 𐴙 KINNA YA
bau	𐴁𐴝𐴗	𐴁 BA, 𐴝 VOWEL SIGN A, 𐴗 KINNA WA
bui	𐴁𐴟𐴙	𐴁 BA, 𐴟 VOWEL SIGN U, 𐴙 KINNA YA
bei	𐴁𐴠𐴙	𐴁 BA, 𐴠 VOWEL SIGN E, 𐴙 KINNA YA
boi	𐴁𐴡𐴙	𐴁 BA, 𐴡 VOWEL SIGN O, 𐴙 KINNA YA
bou	𐴁𐴡𐴗	𐴁 BA, 𐴡 VOWEL SIGN O, 𐴗 KINNA WA

5.5 Nasalization

The 𐴣 NA KHONNA is used for indicating vowel nasalization:

	Display	Input →
baṃ	𐴁𐴝𐴣	𐴁 BA, 𐴝 VOWEL SIGN A, 𐴣 NA KHONNA
baṃb	𐴁𐴝𐴣𐴁	𐴁 BA, 𐴝 VOWEL SIGN A, 𐴣 NA KHONNA, 𐴁 BA
báaṃb	𐴁𐴦𐴝𐴣𐴁	𐴁 BA, 𐴝 VOWEL SIGN A, ◌𐴦 SIGN TANA, 𐴝 VOWEL SIGN A, 𐴣 NA KHONNA, 𐴁 BA

5.6 Consonant gemination

In the input sequence the ◌𐴧 SIGN TOSSI is placed after the letter to be marked as geminated, but before a vowel sign, SAKIN, or NA KHONNA:

	Display	Input →
bba	𐴁𐴧𐴝	𐴁 BA, ◌𐴧 SIGN TOSSI, 𐴝 VOWEL SIGN A
bbaṃ	𐴁𐴧𐴝𐴣	𐴁 BA, ◌𐴧 SIGN TOSSI, 𐴝 VOWEL SIGN A, 𐴣 NA KHONNA
bb	𐴁𐴧𐴢	𐴁 BA, ◌𐴧 SIGN TOSSI, 𐴢 SAKIN
gga	𐴒𐴧𐴝	𐴒 GA, ◌𐴧 SIGN TOSSI, 𐴝 VOWEL SIGN A
ggáa	𐴒𐴧𐴝𐴥	𐴒 BA, ◌𐴧 SIGN TOSSI, 𐴝 VOWEL SIGN A, ◌𐴥 SIGN TELA

In general, the sign TOSSI is centered above a letter, as shown above. However, its graphical placement may vary depending upon the glyphic attributes of the base letter. If the base letter has a high ascender, the TOSSI may be positioned to the left. In such cases, the TOSSI would occur in the same position where a tonal sign would be placed. When both TOSSI and a tonal sign are present, the TOSSI is placed first and the tonal sign to its left. The logical input order does not change.

	tossi pos.	Display	Input →
bbáa	above	𐴁𐴧𐴀	𐴁 BA, ◌𐴧 SIGN TOSSI, 𐴀 VOWEL SIGN A, ◌ SIGN TELA
	left	𐴁𐴧𐴀	𐴁 BA, ◌𐴧 SIGN TOSSI, 𐴀 VOWEL SIGN A, ◌ SIGN TELA

5.7 Digits

As in Arabic, digits in the Hanifi Rohingya are written from left to right:

	Display	Input →
01234	𐴰𐴱𐴲𐴳𐴴	𐴰 ZERO, 𐴱 ONE, 𐴲 TWO, 𐴳 THREE, 𐴴 FOUR
56789	𐴵𐴶𐴷𐴸𐴹	𐴵 FIVE 𐴶 SIX, 𐴷 SEVEN, 𐴸 EIGHT, 𐴹 NINE

The Arabic style ٠ U+0660 ARABIC DIGIT ZERO is attested as a glyphic variant for 𐴰 ZERO.

5.8 Punctuation

There is no script-specific punctuation. The following characters are commonly used for delimiting various text boundaries (see figure 26):

- ، U+060C ARABIC COMMA
- ؛ U+061F ARABIC SEMICOLON
- ؟ U+061F ARABIC QUESTION MARK
- ۔ U+06D4 ARABIC FULL STOP

The available sources show some unconventional usage of ، U+060C ARABIC COMMA. For instance, figure 13 illustrates repetitions of U+060C ARABIC COMMA for indicating quotes, eg. ،،𐴘،، (= “*ya*”). The last line of figure 17 shows another use of ARABIC COMMA for quotations in ، 𐴁𐴡𐴌𐴔𐴀 ، (= ‘Borma’).

The above marks of punctuation should be extended for use with Hanifi Rohingya. Common marks of punctuation include:

. U+002E FULL STOP

« U+00AB LEFT-POINTING DOUBLE ANGLE QUOTATION MARK

» U+00BB RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK

Rounded forms of these guillemets are shown in figure 15, in the heading: «ʻ» 𐴀𐴝𐴑𐴈𐴞𐴁𐴝𐴓 «˜» 𐴈𐴝𐴇𐴝

Both ۔ U+06D4 ARABIC FULL STOP and the regular . U+002E FULL STOP are used commonly. In writing, the width of the full stop may vary from a dot to a dash.

5.9 Elongation

Elongation of the baseline connection between letters is commonly used in Hanifi Rohingya for justification and other spacing-filling purposes (see figure 14). This elongation is similar to the *tatweel* in Arabic. This feature can be implementing in Rohingya by using the ـ U+0640 ARABIC TATWEEL. The ARABIC TATWEEL can be placed after any letter, vowel sign, or NA KHONNA in any position within a word. The elongation should not break across the end of a line.

𐴁𐴝𐴁 <𐴁 BA, 𐴝 VOWEL SIGN A, 𐴁 BA>

𐴁𐴝ـ𐴁 <𐴁 BA, 𐴝 VOWEL SIGN A, ـ ARABIC TATWEEL, 𐴁 BA>

𐴁ـ𐴝𐴁 <𐴁 BA, ـ ARABIC TATWEEL, 𐴝 VOWEL SIGN A, 𐴁 BA>

𐴁𐴝ـ𐴁ـ <𐴁 BA, 𐴝 VOWEL SIGN A, ـ ARABIC TATWEEL, 𐴁 BA, ـ ARABIC TATWEEL>

5.10 Collation

The sort order for Hanifi Rohingya is as follows:

𐴀 A < 𐴝 VOWEL SIGN A < 𐴞 VOWEL SIGN I < 𐴠 VOWEL SIGN U < 𐴟 VOWEL SIGN E < 𐴡 VOWEL SIGN O < 𐴁 BA < 𐴂 PA < 𐴃 TA < 𐴄 TTA < 𐴅 JA < 𐴆 CA < 𐴇 HA < 𐴈 KHA < 𐴉 FA < 𐴊 DA < 𐴋 DDA < 𐴌 RA < 𐴍 RRA < 𐴎 ZA < 𐴏 SHA < 𐴐 SA < 𐴑 KA < 𐴒 GA < 𐴓 LA < 𐴔 MA < 𐴣 NA KHONNA < 𐴕 NA < 𐴖 WA < 𐴗 KINNA WA < 𐴘 YA < 𐴙 KINNA YA < 𐴚 NGA < 𐴛 NYA < 𐴜 VA < 𐴢 SAKIN

6 Character data

6.1 Unicode Character Data

```
10D00;HANIFI ROHINGYA LETTER A;Lo;0;AL;;;;;N;;;;;
10D01;HANIFI ROHINGYA LETTER BA;Lo;0;AL;;;;;N;;;;;
10D02;HANIFI ROHINGYA LETTER PA;Lo;0;AL;;;;;N;;;;;
10D03;HANIFI ROHINGYA LETTER TA;Lo;0;AL;;;;;N;;;;;
10D04;HANIFI ROHINGYA LETTER TTA;Lo;0;AL;;;;;N;;;;;
10D05;HANIFI ROHINGYA LETTER JA;Lo;0;AL;;;;;N;;;;;
10D06;HANIFI ROHINGYA LETTER CA;Lo;0;AL;;;;;N;;;;;
10D07;HANIFI ROHINGYA LETTER HA;Lo;0;AL;;;;;N;;;;;
10D08;HANIFI ROHINGYA LETTER KHA;Lo;0;AL;;;;;N;;;;;
10D09;HANIFI ROHINGYA LETTER FA;Lo;0;AL;;;;;N;;;;;
10D0A;HANIFI ROHINGYA LETTER DA;Lo;0;AL;;;;;N;;;;;
10D0B;HANIFI ROHINGYA LETTER DDA;Lo;0;AL;;;;;N;;;;;
10D0C;HANIFI ROHINGYA LETTER RA;Lo;0;AL;;;;;N;;;;;
10D0D;HANIFI ROHINGYA LETTER RRA;Lo;0;AL;;;;;N;;;;;
10D0E;HANIFI ROHINGYA LETTER ZA;Lo;0;AL;;;;;N;;;;;
10D0F;HANIFI ROHINGYA LETTER SA;Lo;0;AL;;;;;N;;;;;
10D10;HANIFI ROHINGYA LETTER SHA;Lo;0;AL;;;;;N;;;;;
10D11;HANIFI ROHINGYA LETTER KA;Lo;0;AL;;;;;N;;;;;
10D12;HANIFI ROHINGYA LETTER GA;Lo;0;AL;;;;;N;;;;;
10D13;HANIFI ROHINGYA LETTER LA;Lo;0;AL;;;;;N;;;;;
10D14;HANIFI ROHINGYA LETTER MA;Lo;0;AL;;;;;N;;;;;
10D15;HANIFI ROHINGYA LETTER NA;Lo;0;AL;;;;;N;;;;;
10D16;HANIFI ROHINGYA LETTER WA;Lo;0;AL;;;;;N;;;;;
10D17;HANIFI ROHINGYA LETTER KINNA WA;Lo;0;AL;;;;;N;;;;;
10D18;HANIFI ROHINGYA LETTER YA;Lo;0;AL;;;;;N;;;;;
10D19;HANIFI ROHINGYA LETTER KINNA YA;Lo;0;AL;;;;;N;;;;;
10D1A;HANIFI ROHINGYA LETTER NGA;Lo;0;AL;;;;;N;;;;;
10D1B;HANIFI ROHINGYA LETTER NYA;Lo;0;AL;;;;;N;;;;;
10D1C;HANIFI ROHINGYA LETTER VA;Lo;0;AL;;;;;N;;;;;
10D1D;HANIFI ROHINGYA VOWEL SIGN A;Lo;0;AL;;;;;N;;;;;
10D1E;HANIFI ROHINGYA VOWEL SIGN I;Lo;0;AL;;;;;N;;;;;
10D1F;HANIFI ROHINGYA VOWEL SIGN U;Lo;0;AL;;;;;N;;;;;
10D20;HANIFI ROHINGYA VOWEL SIGN E;Lo;0;AL;;;;;N;;;;;
10D21;HANIFI ROHINGYA VOWEL SIGN O;Lo;0;AL;;;;;N;;;;;
10D22;HANIFI ROHINGYA SAKIN;Lo;0;AL;;;;;N;;;;;
10D23;HANIFI ROHINGYA NA KHONNA;Lo;0;AL;;;;;N;;;;;
10D24;HANIFI ROHINGYA SIGN HARBAY;Mn;230;NSM;;;;;N;;;;;
10D25;HANIFI ROHINGYA SIGN TELA;Mn;230;NSM;;;;;N;;;;;
10D26;HANIFI ROHINGYA SIGN TANA;Mn;230;NSM;;;;;N;;;;;
10D27;HANIFI ROHINGYA SIGN TOSSI;Mn;230;NSM;;;;;N;;;;;
10D30;HANIFI ROHINGYA DIGIT ZERO;Nd;0;AN;;0;0;0;N;;;;;
10D31;HANIFI ROHINGYA DIGIT ONE;Nd;0;AN;;1;1;1;N;;;;;
10D32;HANIFI ROHINGYA DIGIT TWO;Nd;0;AN;;2;2;2;N;;;;;
10D33;HANIFI ROHINGYA DIGIT THREE;Nd;0;AN;;3;3;3;N;;;;;
10D34;HANIFI ROHINGYA DIGIT FOUR;Nd;0;AN;;4;4;4;N;;;;;
10D35;HANIFI ROHINGYA DIGIT FIVE;Nd;0;AN;;5;5;5;N;;;;;
10D36;HANIFI ROHINGYA DIGIT SIX;Nd;0;AN;;6;6;6;N;;;;;
10D37;HANIFI ROHINGYA DIGIT SEVEN;Nd;0;AN;;7;7;7;N;;;;;
10D38;HANIFI ROHINGYA DIGIT EIGHT;Nd;0;AN;;8;8;8;N;;;;;
10D39;HANIFI ROHINGYA DIGIT NINE;Nd;0;AN;;9;9;9;N;;;;;
```

6.2 Linebreaking

```
10D00..10D1C;AL  # Lo  [29] HANIFI ROHINGYA LETTER A..HANIFI ROHINGYA LETTER VA
```

```
10D1D..10D21;BA  # Lo  [21] HANIFI ROHINGYA VOWEL SIGN A..HANIFI ROHINGYA VOWEL SIGN O
10D22..10D23;AL  # Lo   [2] HANIFI ROHINGYA SAKIN..HANIFI ROHINGYA NA KHONNA
10D24..10D27;CM  # Mn   [4] HANIFI ROHINGYA SIGN HARBAY..HANIFI ROHINGYA SIGN TOSSI
10D30..10D31;NU  # Nd  [10] HANIFI ROHINGYA DIGIT ZERO..HANIFI ROHINGYA DIGIT NINE
```

6.3 Arabic Shaping Data

```
10D00; HANIFI ROHINGYA A; L; HANIFI ROHINGYA A
10D01; HANIFI ROHINGYA BA; D; HANIFI ROHINGYA BA
10D02; HANIFI ROHINGYA PA; D; HANIFI ROHINGYA PA
10D03; HANIFI ROHINGYA TA; D; HANIFI ROHINGYA TA
10D04; HANIFI ROHINGYA TTA; D; HANIFI ROHINGYA TTA
10D05; HANIFI ROHINGYA JA; D; HANIFI ROHINGYA JA
10D06; HANIFI ROHINGYA CA; D; HANIFI ROHINGYA CA
10D07; HANIFI ROHINGYA HA; D; HANIFI ROHINGYA HA
10D08; HANIFI ROHINGYA KHA; D; HANIFI ROHINGYA KHA
10D09; HANIFI ROHINGYA PA WITH DOT ABOVE; D; HANIFI ROHINGYA PA
10D0A; HANIFI ROHINGYA DA; D; HANIFI ROHINGYA DA
10D0B; HANIFI ROHINGYA DDA; D; HANIFI ROHINGYA DDA
10D0C; HANIFI ROHINGYA RA; D; HANIFI ROHINGYA RA
10D0D; HANIFI ROHINGYA RRA; D; HANIFI ROHINGYA RRA
10D0E; HANIFI ROHINGYA ZA; D; HANIFI ROHINGYA ZA
10D0F; HANIFI ROHINGYA SA; D; HANIFI ROHINGYA SA
10D10; HANIFI ROHINGYA SHA; D; HANIFI ROHINGYA SHA
10D11; HANIFI ROHINGYA KA; D; HANIFI ROHINGYA KA
10D12; HANIFI ROHINGYA GA; D; HANIFI ROHINGYA GA
10D13; HANIFI ROHINGYA LA; D; HANIFI ROHINGYA LA
10D14; HANIFI ROHINGYA MA; D; HANIFI ROHINGYA MA
10D15; HANIFI ROHINGYA NA; D; HANIFI ROHINGYA NA
10D16; HANIFI ROHINGYA WA; D; HANIFI ROHINGYA WA
10D17; HANIFI ROHINGYA KINNA WA; D; HANIFI ROHINGYA KINNA WA
10D18; HANIFI ROHINGYA YA; D; HANIFI ROHINGYA YA
10D19; HANIFI ROHINGYA KINNA YA; D; HANIFI ROHINGYA KINNA YA
10D1A; HANIFI ROHINGYA NGA; D; HANIFI ROHINGYA NGA
10D1B; HANIFI ROHINGYA NYA; D; HANIFI ROHINGYA NYA
10D1C; HANIFI ROHINGYA PA WITH THREE DOTS ABOVE; D; HANIFI ROHINGYA PA
10D1D; HANIFI ROHINGYA VOWEL SIGN A; D; HANIFI ROHINGYA VOWEL SIGN A
10D1E; HANIFI ROHINGYA DOTLESS KINNA YA WITH RIGHT-FACING HOOK BELOW; D;
       HANIFI ROHINGYA KINNA YA
10D1F; HANIFI ROHINGYA VOWEL SIGN U; D; HANIFI ROHINGYA VOWEL SIGN U
10D20; HANIFI ROHINGYA DOTLESS KINNA YA WITH LEFT-FACING HOOK BELOW; D;
       HANIFI ROHINGYA KINNA YA
10D21; HANIFI ROHINGYA VOWEL SIGN O; D; HANIFI ROHINGYA VOWEL SIGN O
10D22; HANIFI ROHINGYA SAKIN; R; HANIFI ROHINGYA SAKIN
10D23; HANIFI ROHINGYA DOTLESS KINNNA YA WITH DOT ABOVE; D;
       HANIFI ROHINGYA KINNA YA
```

6.4 Script Extensions

The following characters should be extended for usage in Rohingya:

```
0640 ; # Lm  ARABIC TATWEEL
060C ; # Po  ARABIC COMMA
061B ; # Po  ARABIC SEMICOLON
061F ; # Po  ARABIC QUESTION MARK
06D4 ; # Po  ARABIC FULL STOP
```

6.5 'Confusable' Characters

Some Rohingya letters resemble those found of Arabic. Attention to these 'confusables' is particularly important because Arabic is commonly used in Rohingya documents.

```
10D03 HANIFI ROHINGYA LETTER PA         ;  0648 ARABIC LETTER WAW
10D07 HANIFI ROHINGYA LETTER HA         ;  06BE ARABIC LETTER HEH DOACHASHMEE
10D08 HANIFI ROHINGYA LETTER KHA        ;  06A9 ARABIC LETTER KEHEH
10D09 HANIFI ROHINGYA LETTER FA         ;  06CF ARABIC LETTER WAW WITH DOT ABOVE
10D0B HANIFI ROHINGYA LETTER DDA        ;  0637 ARABIC LETTER TAH
10D0C HANIFI ROHINGYA LETTER RRA        ;  0637 ARABIC LETTER TAH
10D11 HANIFI ROHINGYA LETTER KA         ;  0637 ARABIC LETTER TAH
10D12 HANIFI ROHINGYA LETTER GA         ;  FECB ARABIC LETTER AIN INITIAL FORM
10D13 HANIFI ROHINGYA LETTER LA         ;  0644 ARABIC LETTER LAM
10D14 HANIFI ROHINGYA LETTER MA         ;  FEE3 ARABIC LETTER MEEM INITIAL FORM
10D19 HANIFI ROHINGYA LETTER KINNA YA   ;  FE91 ARABIC LETTER BEH INITIAL FORM
10D1C HANIFI ROHINGYA LETTER VA         ;  06A8 ARABIC LETTER QAF WITH THREE DOTS ABOVE
10D1D HANIFI ROHINGYA VOWEL SIGN A      ;  FEBB ARABIC LETTER SAD INITIAL FORM
10D23 HANIFI ROHINGYA NA KHONNA         ;  0632 ARABIC LETTER ZAIN
10D23 HANIFI ROHINGYA NA KHONNA         ;  FEE7 ARABIC LETTER NOON INITIAL FORM
10D25 HANIFI ROHINGYA SIGN TELA         ;  0654 ARABIC HAMZA ABOVE
10D26 HANIFI ROHINGYA SIGN TANA         ;  0653 ARABIC MADDAH ABOVE
10D27 HANIFI ROHINGYA SIGN TOSSI        ;  0651 ARABIC SHADDA
```

7 References

Anderson, Deborah; et. al. 2016a. "Recommendations to UTC #146 January 2016 on Script Proposals". L2/16-037. http://www.unicode.org/L2/L2016/16037-script-rec.pdf

———. 2017. "Recommendations to UTC #150 January 2017 on Script Proposals". L2/17-037. http://www.unicode.org/L2/L2017/17037-script-ad-hoc.pdf

Noor, Mohammad. 2011. "Rohingya mother Language (letters) Part 1.flv". http://www.youtube.com/watch?v=w4h6w6NyvOU

Pandey, Anshuman. 2012. "Preliminary Proposal to Encode the Rohingya Script". L2/12-214. http://www.unicode.org/L2/L2012/12214-rohingya.pdf

———. 2015. "Proposal to encode the Hanifi Rohingya script in Unicode". L2/15-278R. http://www.unicode.org/L2/L2015/15278r-hanifi-rohingya.pdf

Priest, Lorna A.; Hosken, Martin. 2010. "Proposal to add Arabic script characters for African and Asian languages". L2/10-288R. http://www.unicode.org/L2/L2010/10288r-arabic-proposal.pdf

Rohingya Education Board Myanmar. *Ruhainggya Zubanor Fonna: Hisab* ["Script of the Rohingya language: Counting"]. Ek kelasottu dui kelas: lego ar foro ["From Class 1 to Class 2: Read and Write"].

Rohingya Language Committee. [A]. *Kayda Ruwainggya Zubanor* [Primer of the Rohingya Language].

———. [B]. *Ruwaingya Zubanor Foyla Kitab* [First Book of the Rohingya Language].

Wikipedia. “Rohingya language’.
`https://en.wikipedia.org/wiki/Rohingya_language`

8 Acknowledgments

I extremely grateful to Muhammad Noor, who has provided me with numerous source materials and who has patiently answered my questions about the script over the years. I also thank the Rohingya Language Committee for responding to my questions. I am thankful to Mattias Persson and Ian James for bringing the Rohingya script to my attention some years ago. My efforts on this work would not have been possible without Lorna Priest, who introduced me to James Lloyd-Williams, who in turn provided copies of Rohingya primers.

The present proposal was funded in part by the Adopt-A-Character Program of the Unicode Consortium. A previous version was made possible in part through a Google Research Award, granted to Deborah Anderson for the Script Encoding Initiative, which funded a post-doctoral research position for me in the Department of Linguistics, University of California, Berkeley during 2015–2016. Preliminary research was made possible through the Script Encoding Initiative at Berkeley. Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the Unicode Consortium; the University of California, Berkeley; or Google.

	10D0	10D1	10D2	10D3
0	𐴀 10D00	𐴐 10D10	𐴠 10D20	𐴰 10D30
1	𐴁 10D01	𐴑 10D11	𐴡 10D21	𐴱 10D31
2	𐴂 10D02	𐴒 10D12	𐴢 10D22	𐴲 10D32
3	𐴃 10D03	𐴓 10D13	𐴣 10D23	𐴳 10D33
4	𐴄 10D04	𐴔 10D14	◌𐴤 10D24	𐴴 10D34
5	𐴅 10D05	𐴕 10D15	◌𐴥 10D25	𐴵 10D35
6	𐴆 10D06	𐴖 10D16	◌𐴦 10D26	𐴶 10D36
7	𐴇 10D07	𐴗 10D17	◌𐴧 10D27	𐴷 10D37
8	𐴈 10D08	𐴘 10D18		𐴸 10D38
9	𐴉 10D09	𐴙 10D19		𐴹 10D39
A	𐴊 10D0A	𐴚 10D1A		
B	𐴋 10D0B	𐴛 10D1B		
C	𐴌 10D0C	𐴜 10D1C		
D	𐴍 10D0D	𐴝 10D1D		
E	𐴎 10D0E	𐴞 10D1E		
F	𐴏 10D0F	𐴟 10D1F		

Consonants

10D00 𐴀 HANIFI ROHINGYA LETTER A
10D01 𐴁 HANIFI ROHINGYA LETTER BA
10D02 𐴂 HANIFI ROHINGYA LETTER PA
10D03 𐴃 HANIFI ROHINGYA LETTER TA
10D04 𐴄 HANIFI ROHINGYA LETTER TTA
10D05 𐴅 HANIFI ROHINGYA LETTER JA
10D06 𐴆 HANIFI ROHINGYA LETTER CA
10D07 𐴇 HANIFI ROHINGYA LETTER HA
10D08 𐴈 HANIFI ROHINGYA LETTER KHA
10D09 𐴉 HANIFI ROHINGYA LETTER FA
10D0A 𐴊 HANIFI ROHINGYA LETTER DA
10D0B 𐴋 HANIFI ROHINGYA LETTER DDA
10D0C 𐴌 HANIFI ROHINGYA LETTER RA
10D0D 𐴍 HANIFI ROHINGYA LETTER RRA
10D0E 𐴎 HANIFI ROHINGYA LETTER ZA
10D0F 𐴏 HANIFI ROHINGYA LETTER SA
10D10 𐴐 HANIFI ROHINGYA LETTER SHA
10D11 𐴑 HANIFI ROHINGYA LETTER KA
10D12 𐴒 HANIFI ROHINGYA LETTER GA
10D13 𐴓 HANIFI ROHINGYA LETTER LA
10D14 𐴔 HANIFI ROHINGYA LETTER MA
10D15 𐴕 HANIFI ROHINGYA LETTER NA
10D16 𐴖 HANIFI ROHINGYA LETTER WA
10D17 𐴗 HANIFI ROHINGYA LETTER KINNA WA
10D18 𐴘 HANIFI ROHINGYA LETTER YA
10D19 𐴙 HANIFI ROHINGYA LETTER KINNA YA
10D1A 𐴚 HANIFI ROHINGYA LETTER NGA
= gan
10D1B 𐴛 HANIFI ROHINGYA LETTER NYA
= nayya

Additional letter

10D1C 𐴜 HANIFI ROHINGYA LETTER VA
• not used at present

Vowel signs

10D1D 𐴝 HANIFI ROHINGYA VOWEL SIGN A
= a-for
10D1E 𐴞 HANIFI ROHINGYA VOWEL SIGN I
= i-for
10D1F 𐴟 HANIFI ROHINGYA VOWEL SIGN U
= u-for
10D20 𐴠 HANIFI ROHINGYA VOWEL SIGN E
= e-for
10D21 𐴡 HANIFI ROHINGYA VOWEL SIGN O
= o-for

Vowel silencer

10D22 𐴢 HANIFI ROHINGYA SAKIN

Nasalization mark

10D23 𐴣 HANIFI ROHINGYA NA KHANNA

Tonal signs

10D24 ◌𐴤 HANIFI ROHINGYA SIGN HARBAY
• short high tone
10D25 ◌𐴥 HANIFI ROHINGYA SIGN TELA
• long rising tone
10D26 ◌𐴦 HANIFI ROHINGYA SIGN TANA
• long falling tone

Gemination sign

10D27 ◌𐴧 HANIFI ROHINGYA SIGN TOSSI

Digits

10D30 𐴰 HANIFI ROHINGYA DIGIT ZERO
10D31 𐴱 HANIFI ROHINGYA DIGIT ONE
10D32 𐴲 HANIFI ROHINGYA DIGIT TWO
10D33 𐴳 HANIFI ROHINGYA DIGIT THREE
10D34 𐴴 HANIFI ROHINGYA DIGIT FOUR
10D35 𐴵 HANIFI ROHINGYA DIGIT FIVE
10D36 𐴶 HANIFI ROHINGYA DIGIT SIX
10D37 𐴷 HANIFI ROHINGYA DIGIT SEVEN
10D38 𐴸 HANIFI ROHINGYA DIGIT EIGHT
10D39 𐴹 HANIFI ROHINGYA DIGIT NINE

	Proposed character name	Fig. 6	Fig. 7	Fig. 8
𐴀	HANIFI ROHINGYA LETTER A	Aa	AA	AA
𐴁	HANIFI ROHINGYA LETTER BA	Ba	BA	BA
𐴂	HANIFI ROHINGYA LETTER PA	Pa	PA	PA
𐴃	HANIFI ROHINGYA LETTER TA	Ta	TA	THA
𐴄	HANIFI ROHINGYA LETTER TTA	Tda	TA	TA
𐴅	HANIFI ROHINGYA LETTER JA	Ja	JA	JA
𐴆	HANIFI ROHINGYA LETTER CA	Cha	CHA	CHA
𐴇	HANIFI ROHINGYA LETTER HA	Ha	HA	HA
𐴈	HANIFI ROHINGYA LETTER KHA	Kha	KHA	KHA
𐴉	HANIFI ROHINGYA LETTER FA	Pha	FA	FA
𐴊	HANIFI ROHINGYA LETTER DA	Da	THA	DHA
𐴋	HANIFI ROHINGYA LETTER DDA	Dha	DA	DA
𐴌	HANIFI ROHINGYA LETTER RA	Ra	RA	RA
𐴍	HANIFI ROHINGYA LETTER RRA	Rda	R̄A	RDA
𐴎	HANIFI ROHINGYA LETTER ZA	Za	ZA	ZA
𐴏	HANIFI ROHINGYA LETTER SA	Sa	SA	SA
𐴐	HANIFI ROHINGYA LETTER SHA	Sha	SHA	SHA
𐴑	HANIFI ROHINGYA LETTER KA	Ka	KA	KA
𐴒	HANIFI ROHINGYA LETTER GA	Ga	GA	GHA
𐴓	HANIFI ROHINGYA LETTER LA	La	LA	LA
𐴔	HANIFI ROHINGYA LETTER MA	Ma	MA	MA
𐴕	HANIFI ROHINGYA LETTER NA	Na	NA	NA
𐴖	HANIFI ROHINGYA LETTER WA	Wa	WA	WA
𐴗	HANIFI ROHINGYA LETTER KINNA WA	Kinnawa	KINNA WA	KINNAWA
𐴘	HANIFI ROHINGYA LETTER YA	Ya	YA	YA
𐴙	HANIFI ROHINGYA LETTER KINNA YA	Kinnaya	KINNA YA	KINNAYA
𐴚	HANIFI ROHINGYA LETTER NGA	Nga	GAN	GHAN
𐴛	HANIFI ROHINGYA LETTER NYA	Naiya	NAYYA	NAYA

Table 1: Romanized names for Rohingya letters as given in various sources.

	Proposed character name	Fig. 6	Fig. 7	Fig. 8
𐴝	HANIFI ROHINGYA VOWEL SIGN A	Aafor	AFOR	AFOR
𐴞	HANIFI ROHINGYA VOWEL SIGN I	Efor	[I]FOR	EIFOR
𐴟	HANIFI ROHINGYA VOWEL SIGN U	Ufor	UFOR	UOFOR
𐴠	HANIFI ROHINGYA VOWEL SIGN E	Aefor	EFOR	EAFOR
𐴡	HANIFI ROHINGYA VOWEL SIGN O	Ofor	OFOR	WOFOR
𐴢	HANIFI ROHINGYA SAKIN	—	—	—
𐴣	HANIFI ROHINGYA NA KHONNA	Nakhonna	NAKONNA	NAKHANNA
◌𐴤	HANIFI ROHINGYA SIGN HARBAY	Haarbai	HARBAY	HAR BAY
◌𐴥	HANIFI ROHINGYA SIGN TELA	Tela	TELA	TAHA LA
◌𐴦	HANIFI ROHINGYA SIGN TANA	Tana	TANA	TA NA
◌𐴧	HANIFI ROHINGYA SIGN TOSSI	Tossi	TASSI	TASSI

Table 2: Romanized names for Rohingya marks as given in various sources.

	Hanifi Rohingya		Arabic
𐴀	HANIFI ROHINGYA LETTER A	ا	U+0627 ARABIC LETTER ALEF
𐴁	HANIFI ROHINGYA LETTER BA	ب	U+0628 ARABIC LETTER BEH
𐴂	HANIFI ROHINGYA LETTER PA	پ	U+067E ARABIC LETTER PEH
𐴃	HANIFI ROHINGYA LETTER TA	ت	U+062A ARABIC LETTER TEH
		ط	U+0637 ARABIC LETTER TAH
𐴄	HANIFI ROHINGYA LETTER TTA	ٹ	U+0679 ARABIC LETTER TTEH
𐴅	HANIFI ROHINGYA LETTER JA	ج	U+062C ARABIC LETTER JEEM
𐴆	HANIFI ROHINGYA LETTER CA	چ	U+0686 ARABIC LETTER TCHEH
𐴇	HANIFI ROHINGYA LETTER HA	ه	U+0647 ARABIC LETTER HE
		ح	U+062D ARABIC LETTER HAH
𐴈	HANIFI ROHINGYA LETTER KHA	خ	U+062E ARABIC LETTER KHAH
𐴉	HANIFI ROHINGYA LETTER FA	ف	U+0641 ARABIC LETTER FEH
𐴊	HANIFI ROHINGYA LETTER DA	د	U+062F ARABIC LETTER DAL
𐴋	HANIFI ROHINGYA LETTER DDA	ڈ	U+0688 ARABIC LETTER DDAL
𐴌	HANIFI ROHINGYA LETTER RA	ر	U+0631 ARABIC LETTER REH
𐴍	HANIFI ROHINGYA LETTER RRA	ڑ	U+0691 ARABIC LETTER RREH
𐴎	HANIFI ROHINGYA LETTER ZA	ز	U+0632 ARABIC LETTER ZAIN
		ض	U+0636 ARABIC LETTER DAD
		ظ	U+0638 ARABIC LETTER ZAH

Table 3: Correspondence between Hanifi Rohingya and Arabic letters (continued in table 4).

	Hanifi Rohingya		Arabic
𐴏	HANIFI ROHINGYA LETTER SA	س	U+0633 ARABIC LETTER SEEN
		ص	U+0635 ARABIC LETTER SAD
𐴐	HANIFI ROHINGYA LETTER SHA	ش	U+0634 ARABIC LETTER SHEEN
𐴑	HANIFI ROHINGYA LETTER KA	ك	U+0643 ARABIC LETTER KAF
		ق	U+0642 ARABIC LETTER QAF
𐴒	HANIFI ROHINGYA LETTER GA	گ	U+06AF ARABIC LETTER GAF
		غ	U+063A ARABIC LETTER GHAIN
𐴓	HANIFI ROHINGYA LETTER LA	ل	U+0644 ARABIC LETTER LAM
𐴔	HANIFI ROHINGYA LETTER MA	م	U+0645 ARABIC LETTER MEEM
𐴕	HANIFI ROHINGYA LETTER NA	ن	U+0646 ARABIC LETTER NOON
𐴖	HANIFI ROHINGYA LETTER WA	و	U+0648 ARABIC LETTER WAW
𐴗	HANIFI ROHINGYA LETTER SMALL WA	و	U+0648 ARABIC LETTER WAW
𐴘	HANIFI ROHINGYA LETTER YA	ي	U+064A ARABIC LETTER YEH
𐴙	HANIFI ROHINGYA LETTER SMALL YA	ي	U+064A ARABIC LETTER YEH
𐴚	HANIFI ROHINGYA LETTER NGA	ڠ	U+06A0 ARABIC LETTER AIN WITH THREE DOTS ABOVE
𐴛	HANIFI ROHINGYA LETTER NYA		—
𐴜	HANIFI ROHINGYA LETTER VA		—

Table 4: Correspondence between Hanifi Rohingya and Arabic letters (continued from table 3).

		Bilabial	Labio-dental	Dental/ Alveolar	Retroflex	Palatal	Velar	Glottal
Plosive	voiceless	p		t̪	ʈ	c	k	ʔ
	voiced	b		d̪	ɖ	ɟ	g	
Nasal		m		n		ɲ	ŋ	
Fricative	voiceless		f	s		ʃ	x	h
	voiced			z				
Flap				ɾ	ɽ			
Approximant	central	w				j		
	lateral			l				

Consonants

	Front	Back
Close	i	u
Close-mid	e	o
Open-mid		ɔ
Open	a	

Vowels

Figure 1: Rohingya phonology (from Wikipedia: “Rohingya language”).

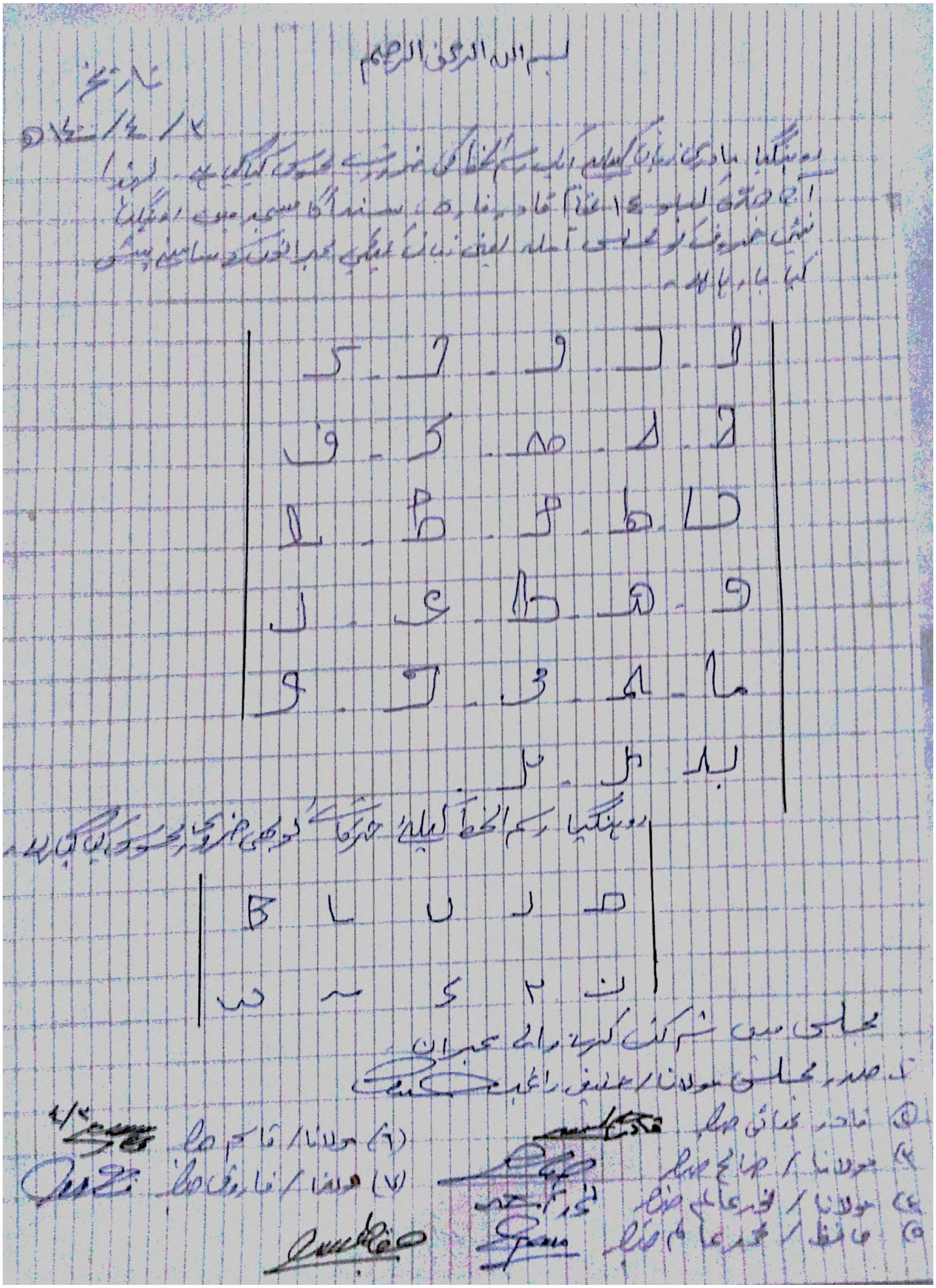


Figure 2: A document showing the Rohingya script as finalized by Maulana Mohammad Hanif and others on February 19, 1980, with signatures of the script committee.

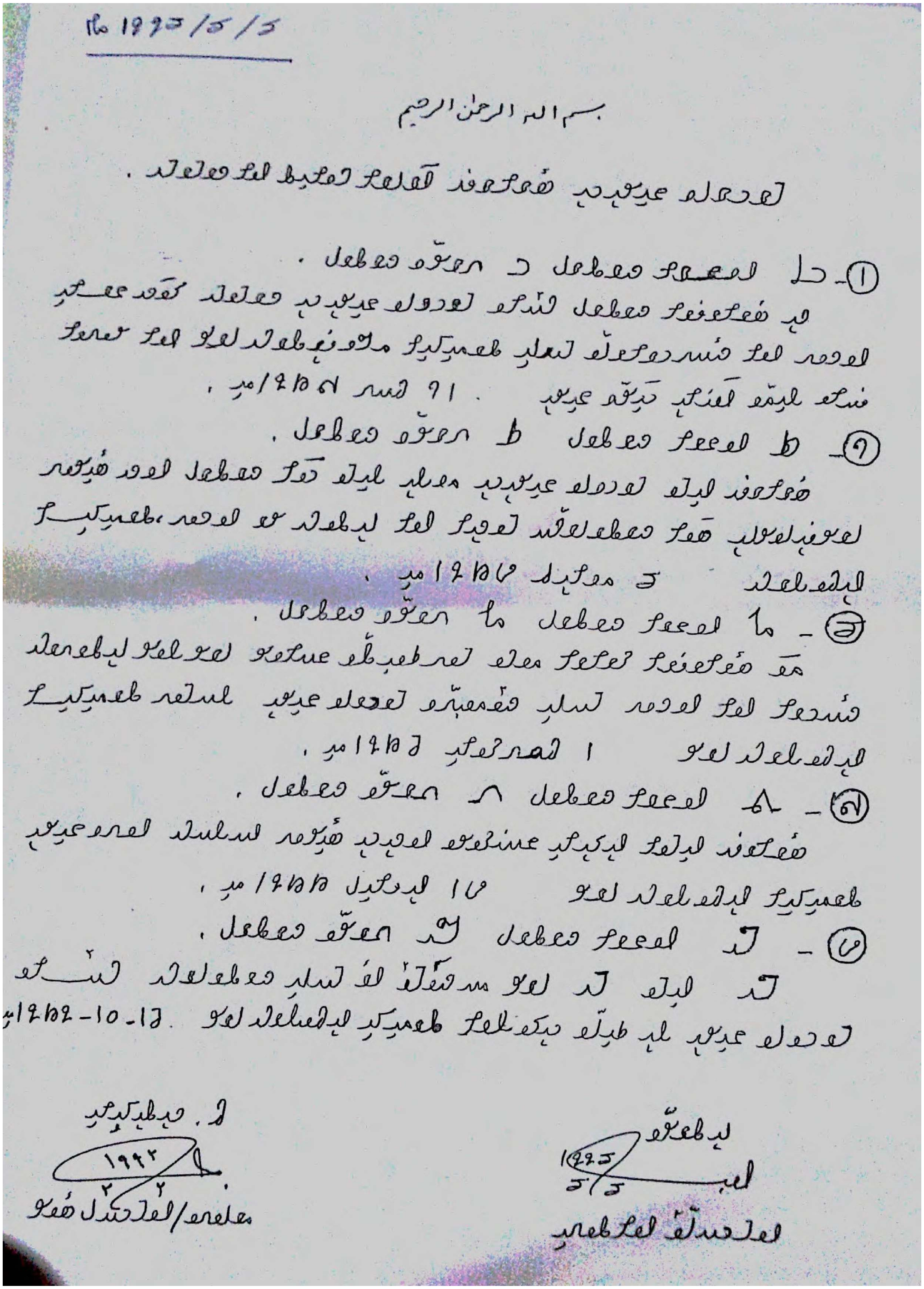
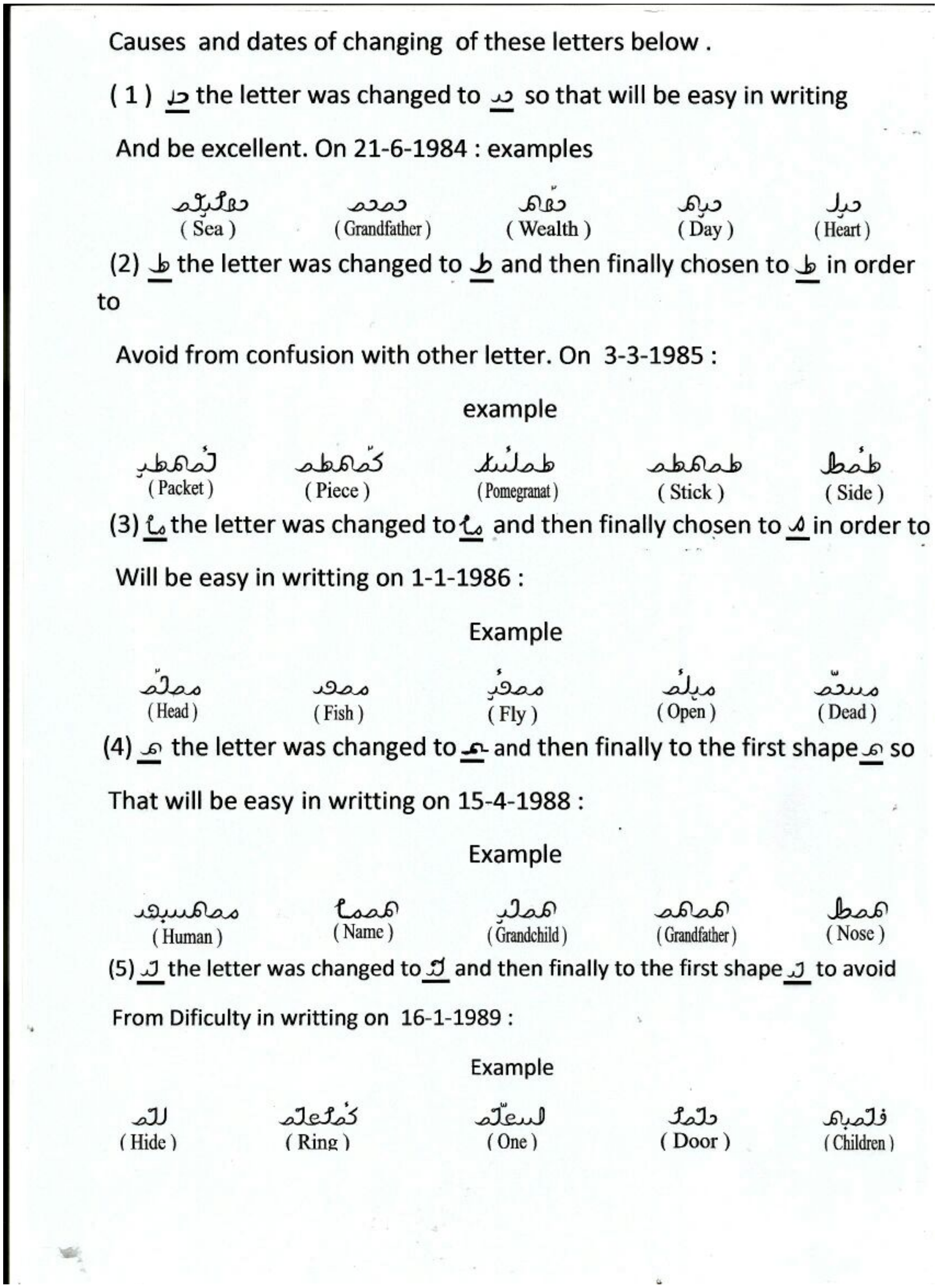


Figure 3: Description of some characters attached to the chart shown in figure 2. A translation is given in figure 4.



Causes and dates of changing of these letters below .

(1) the letter was changed to so that will be easy in writing

And be excellent. On 21-6-1984 : examples

(Sea) (Grandfather) (Wealth) (Day) (Heart)

(2) the letter was changed to and then finally chosen to in order to

Avoid from confusion with other letter. On 3-3-1985 :

example

(Packet) (Piece) (Pomegranat) (Stick) (Side)

(3) the letter was changed to and then finally chosen to in order to

Will be easy in writting on 1-1-1986 :

Example

(Head) (Fish) (Fly) (Open) (Dead)

(4) the letter was changed to and then finally to the first shape so

That will be easy in writting on 15-4-1988 :

Example

(Human) (Name) (Grandchild) (Grandfather) (Nose)

(5) the letter was changed to and then finally to the first shape to avoid

From Dificulty in writting on 16-1-1989 :

Example

(Hide) (Ring) (One) (Door) (Children)

Figure 4: Translation of the text shown in figure 3.

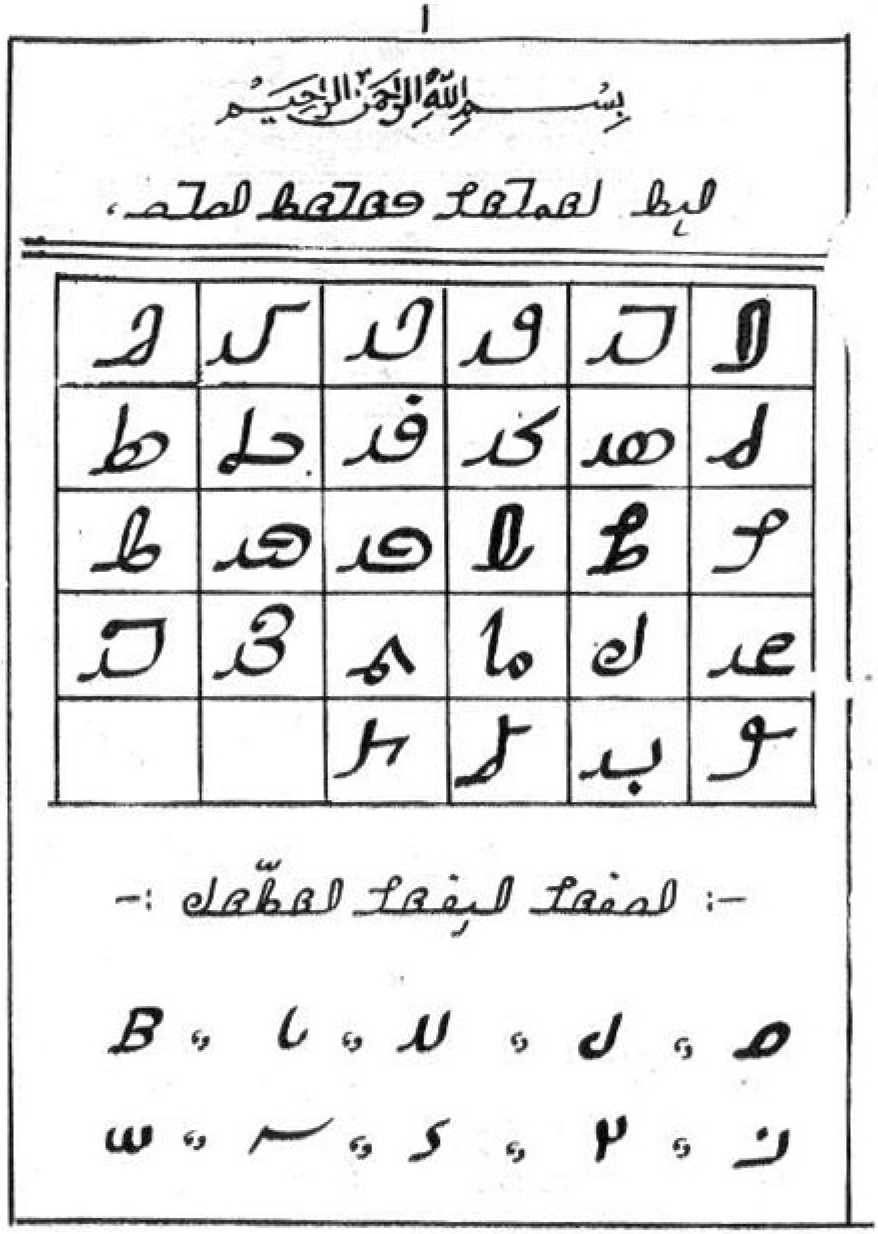


Figure 5: Chart from a hand-written primer (from Ruwainggya Zuban Komiti (A): 1).

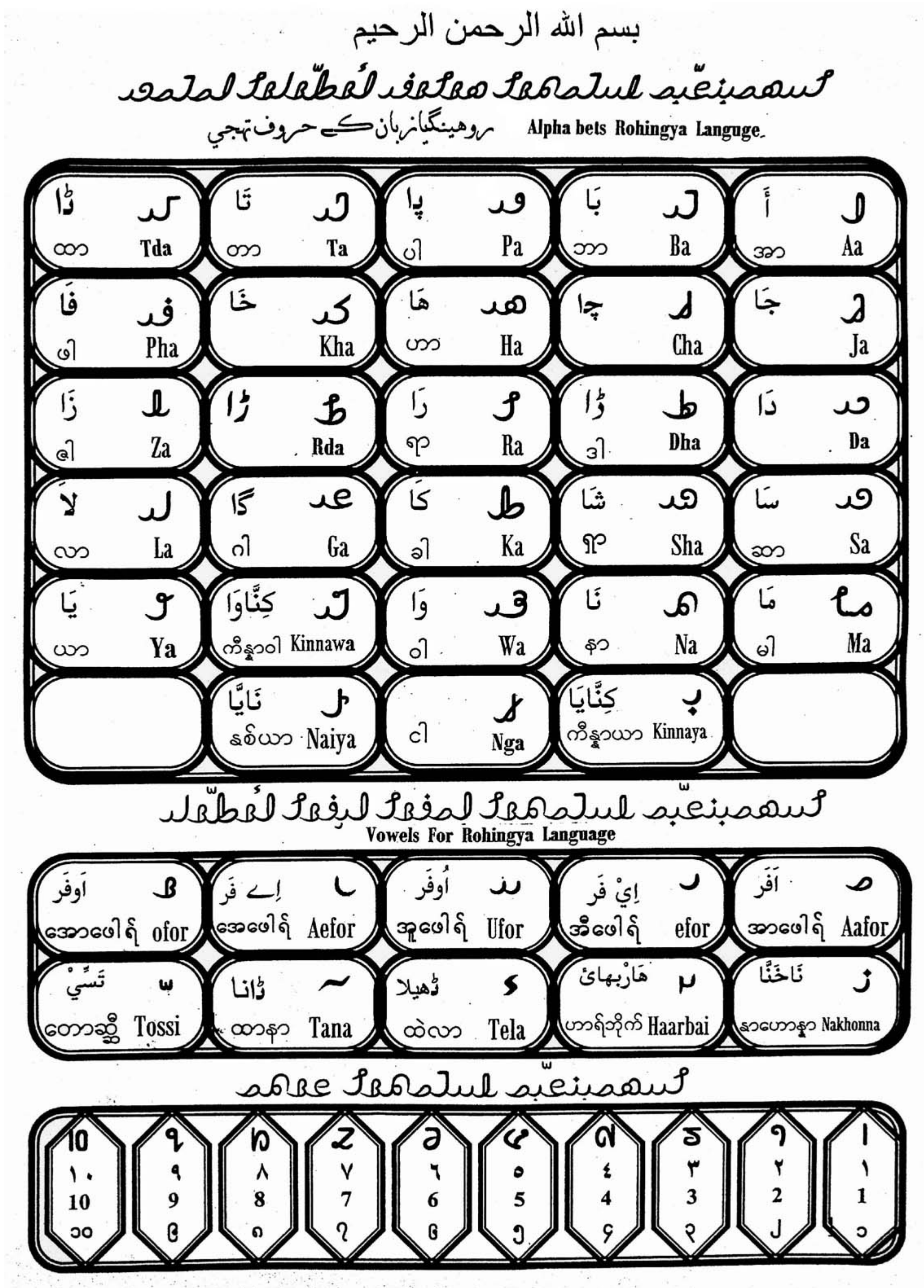


Figure 6: Chart showing Rohingya letters and signs with Urdu and Burmese names.

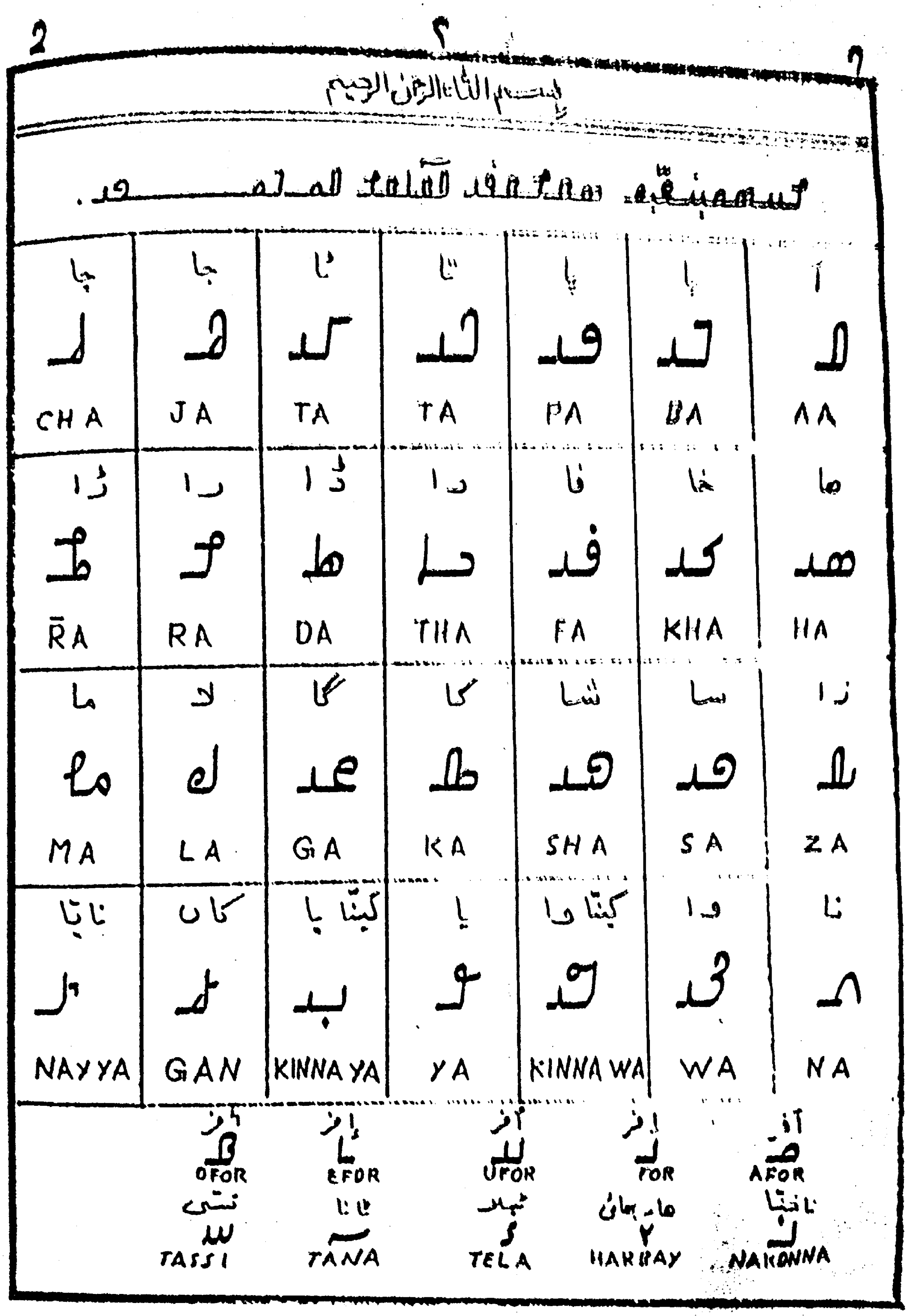


Figure 7: Chart showing Rohingya letters with Arabic correspondences and names in Latin transliteration.

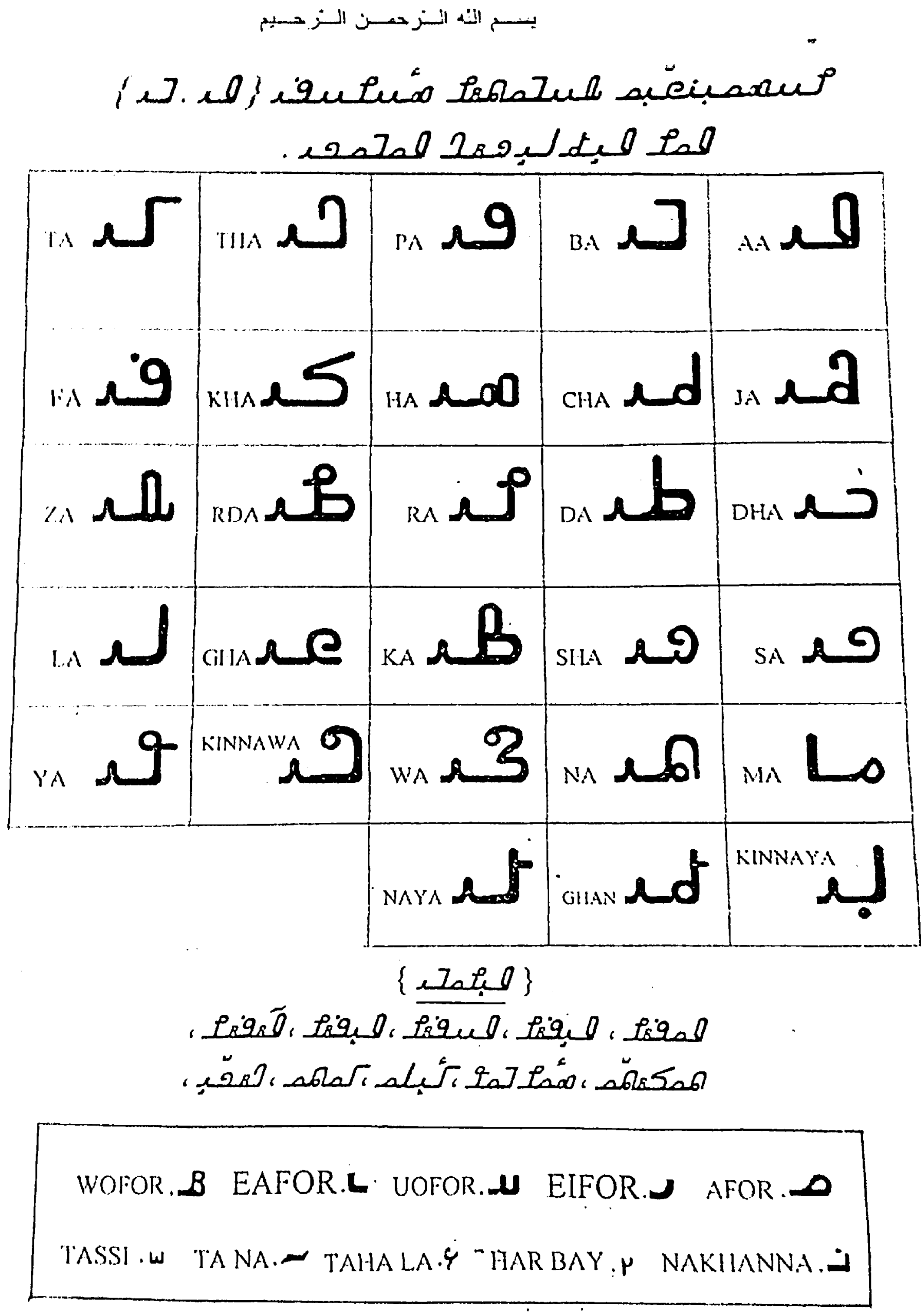


Figure 8: Chart showing Rohingya letters with Latin transliteration.

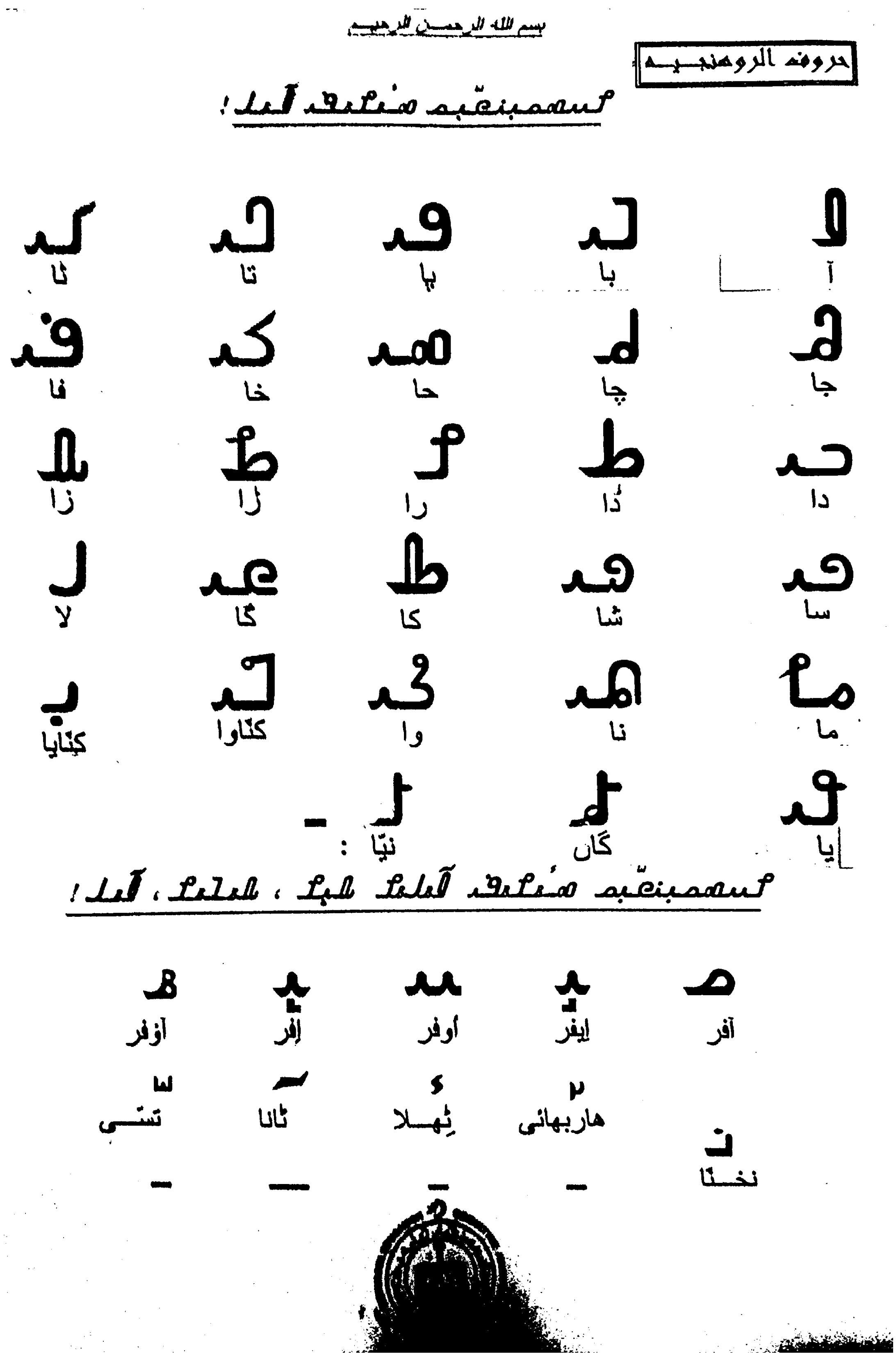


Figure 9: Page from a Rohingya primer showing Rohingya letters with names in Arabic transliteration.

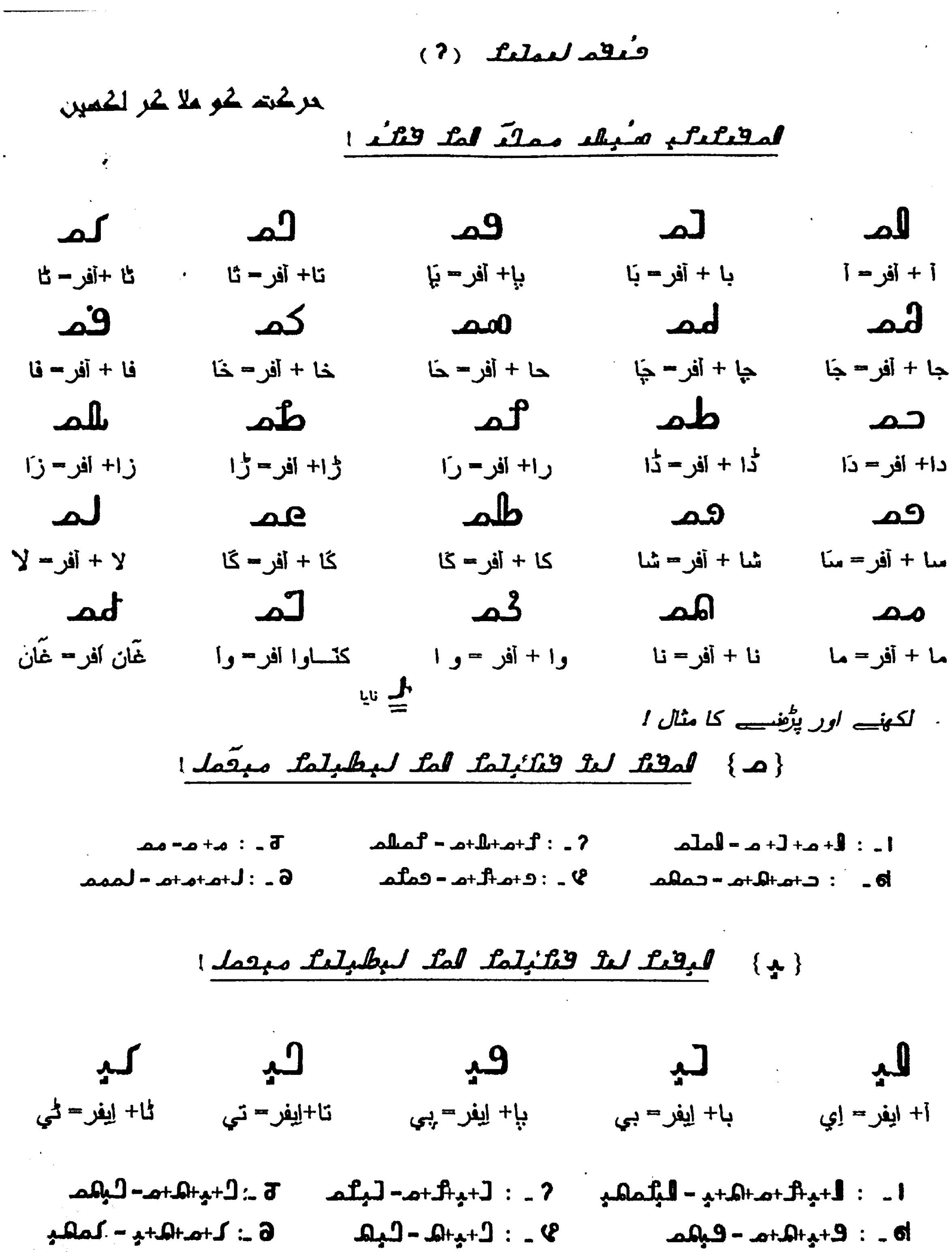


Figure 10: Page from a Rohingya primer showing the method of writing 𐴞 VOWEL SIGN A and 𐴟 VOWEL SIGN I.



Figure 11: Page from a Rohingya primer showing the method of writing VOWEL SIGN U, VOWEL SIGN E, VOWEL SIGN O.



Figure 12: Page from a Rohingya primer showing the method of writing the 𐴣 NA KHONNA, ◌𐴤 SIGN HARBAY, ◌𐴥 SIGN TELA, ◌𐴦 SIGN TANA, ◌𐴧 SIGN TOSSI.

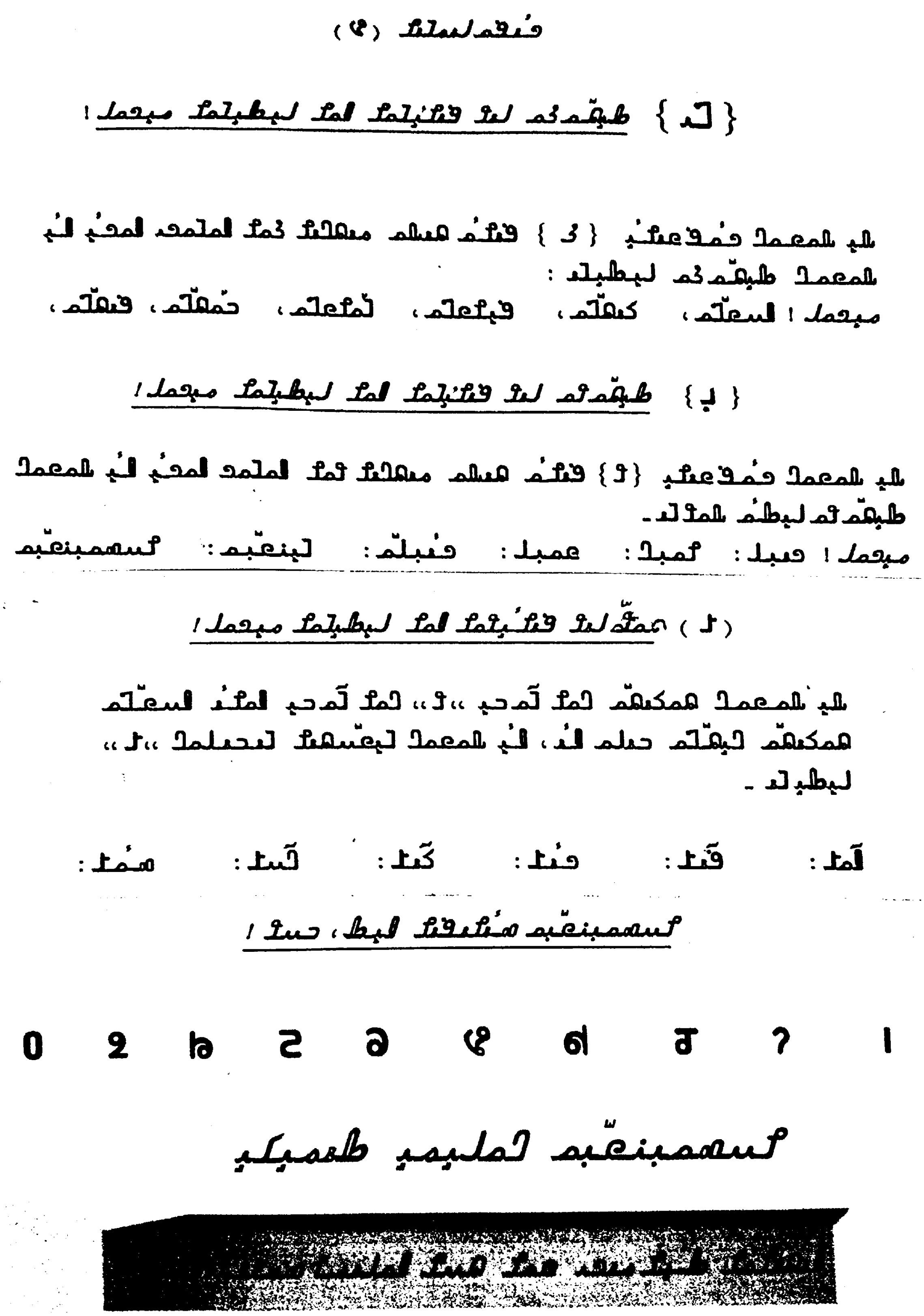


Figure 13: Page from a Rohingya primer describing the use of 𐴃 KINNA WA and 𐴘 KINNA YA. Note the usage of repetitions of ، U+060C ARABIC COMMA for indicating quotes, eg. ،،𐴃،،.

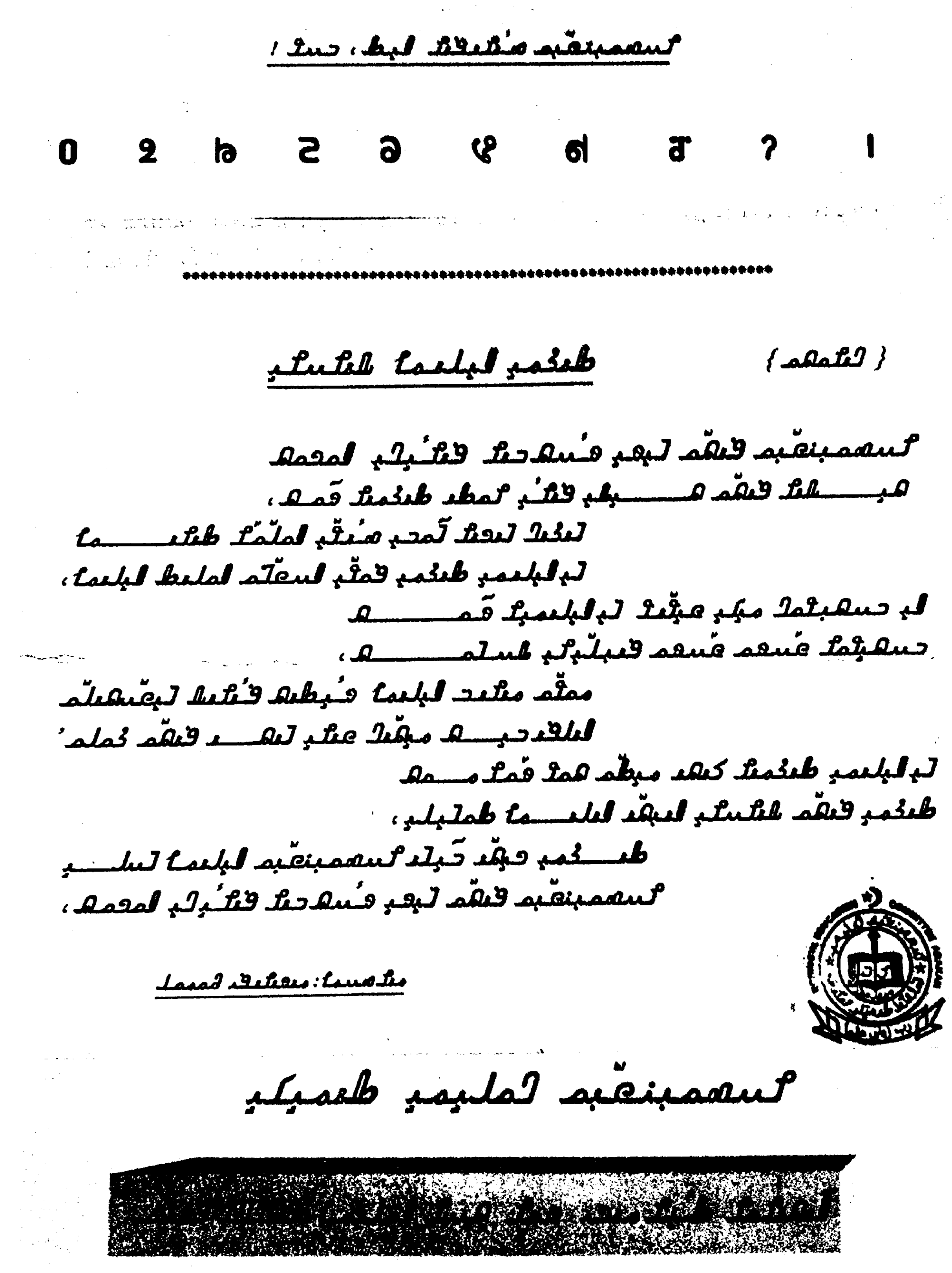


Figure 14: Page from a Rohingya primer showing the use of *tatweel*.

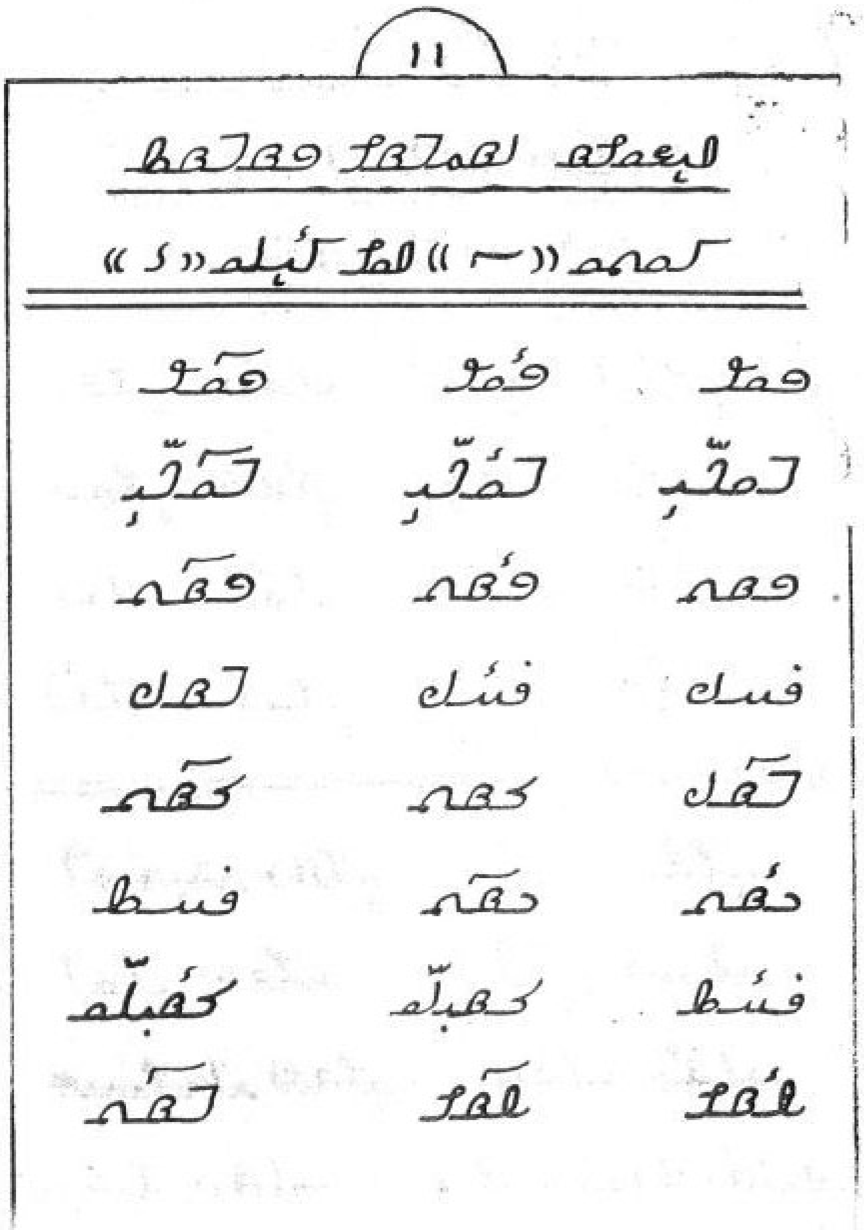


Figure 15: Table showing use of tonal signs from a hand-written primer (from Ruwainggya Zuban Komiti (A): 11).

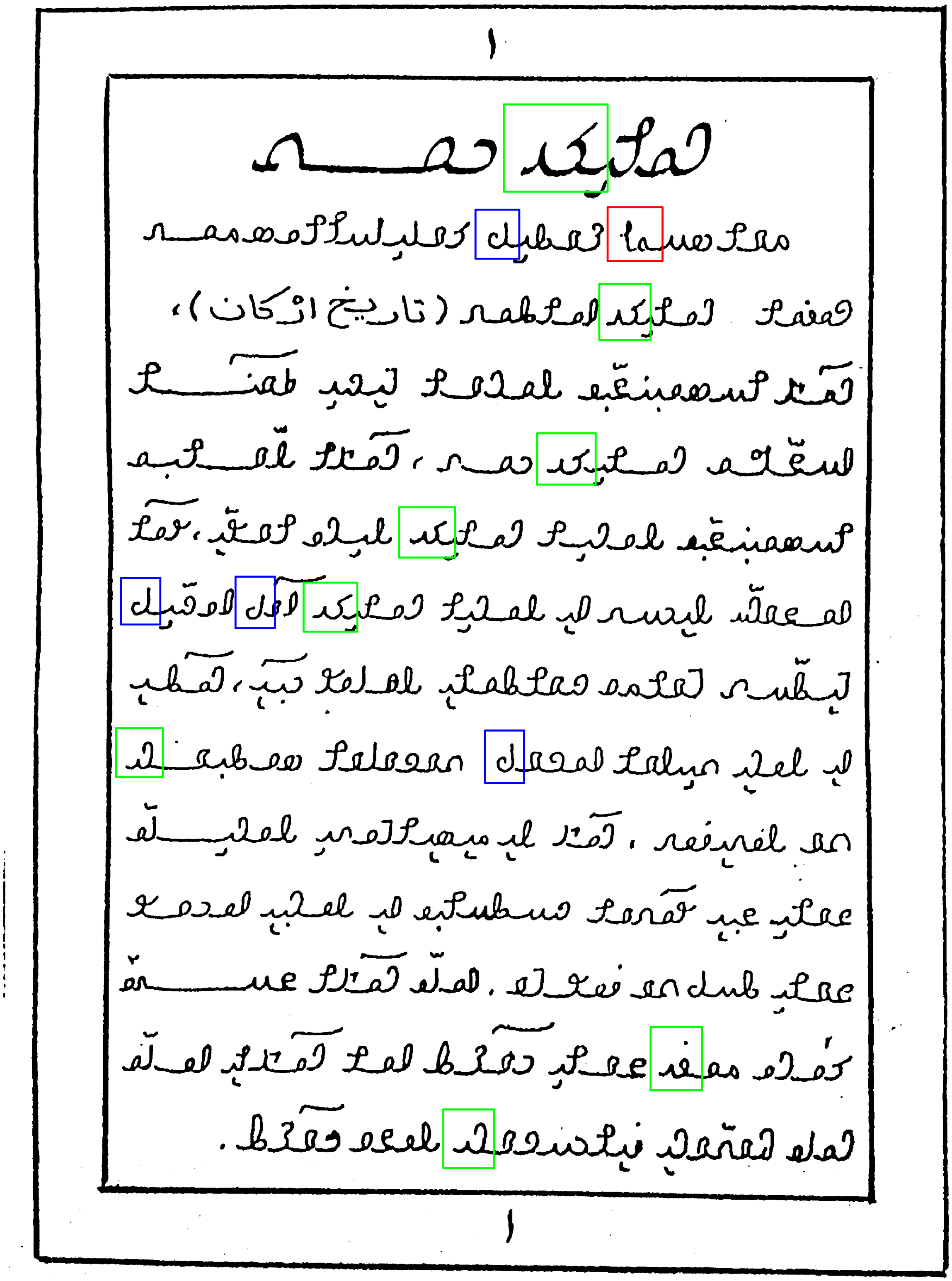


Figure 16: Hand-writen first page of *History of Arakan*. Note the use of the final forms of 𐴔 MA (red) and 𐴓 LA (blue), as well as termination of some words using 𐴢 SAKIN.

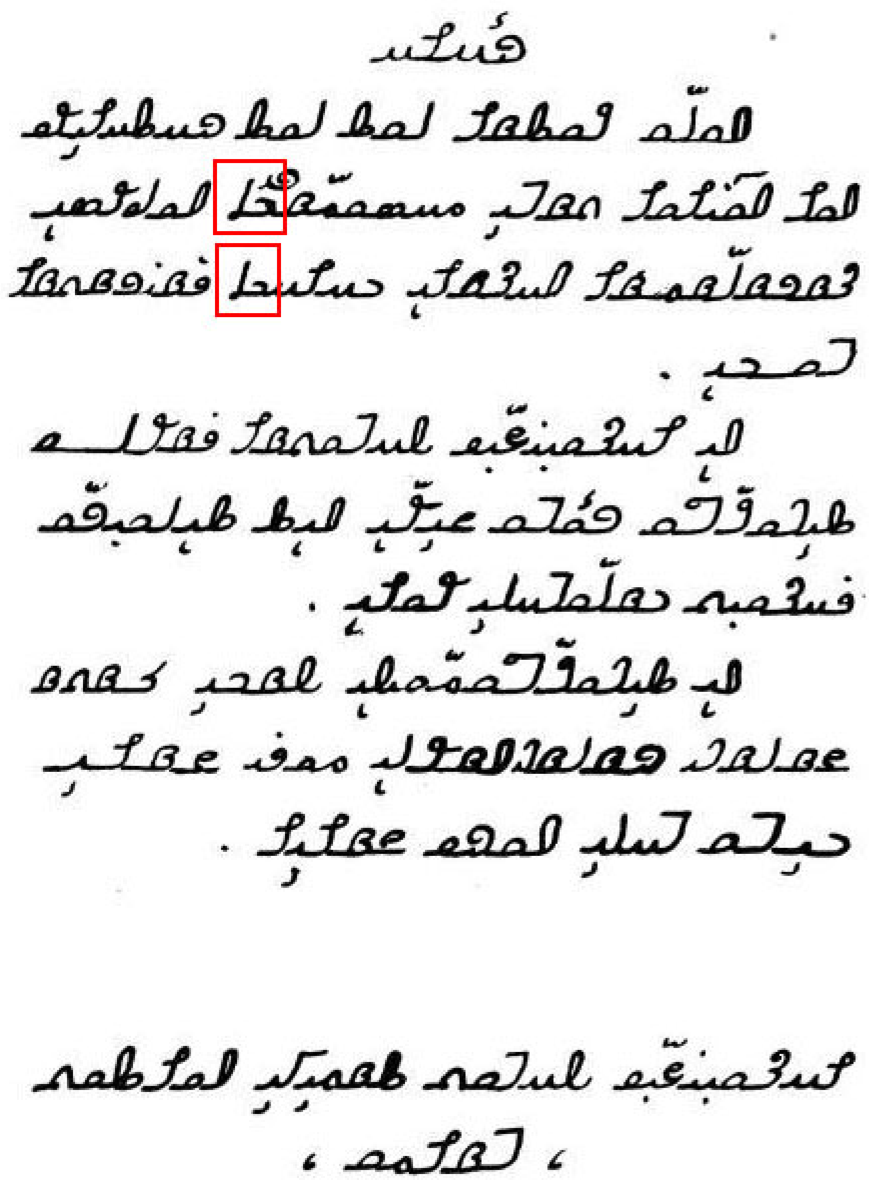


Figure 17: Usage of 𐴊𐴧, an older form of 𐴊𐴧 <𐴊 DA, 𐴧 SAKIN>. (from Ruwainggya Zuban Komiti (B): 1).

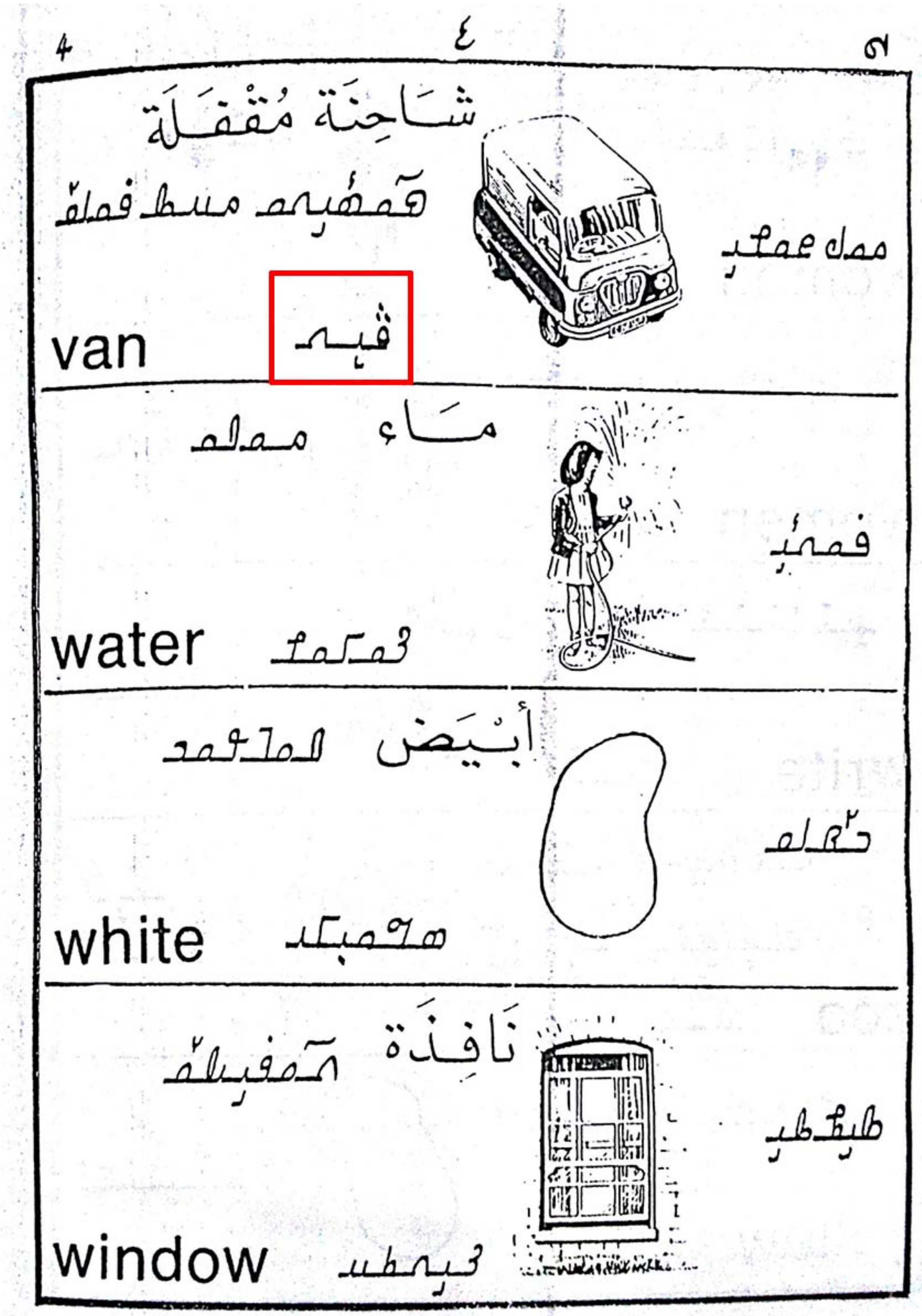


Figure 18: Usage of the letter ڤ *va* in a primer school book for transcribing the English word ‘van’ [illegible]. This character is no longer used in contemporary orthography and has been replaced by 𐴋 WA.

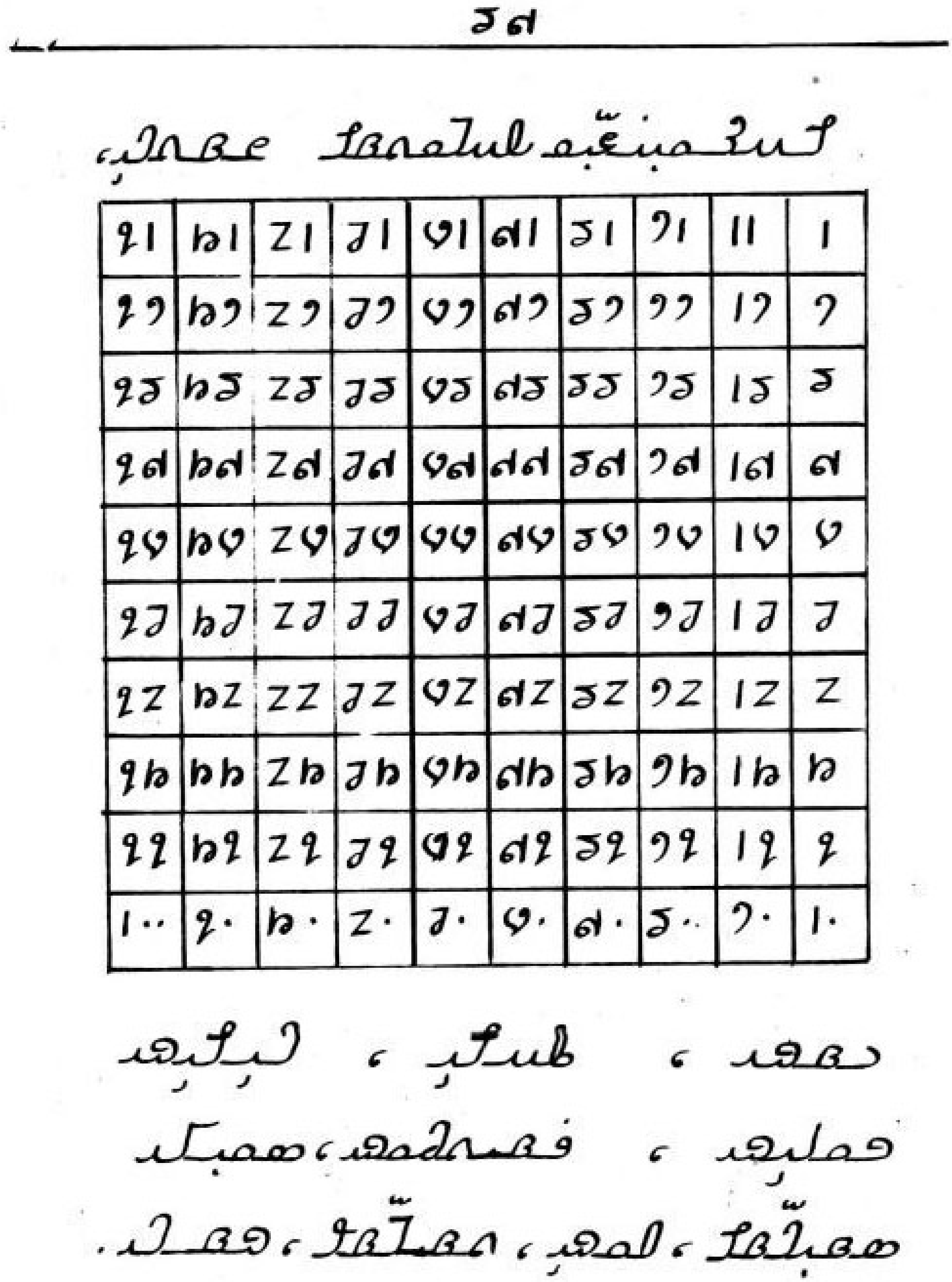


Figure 19: Chart of digits from a hand-written primer (from Ruwainggya Zuban Komiti (B): 34).

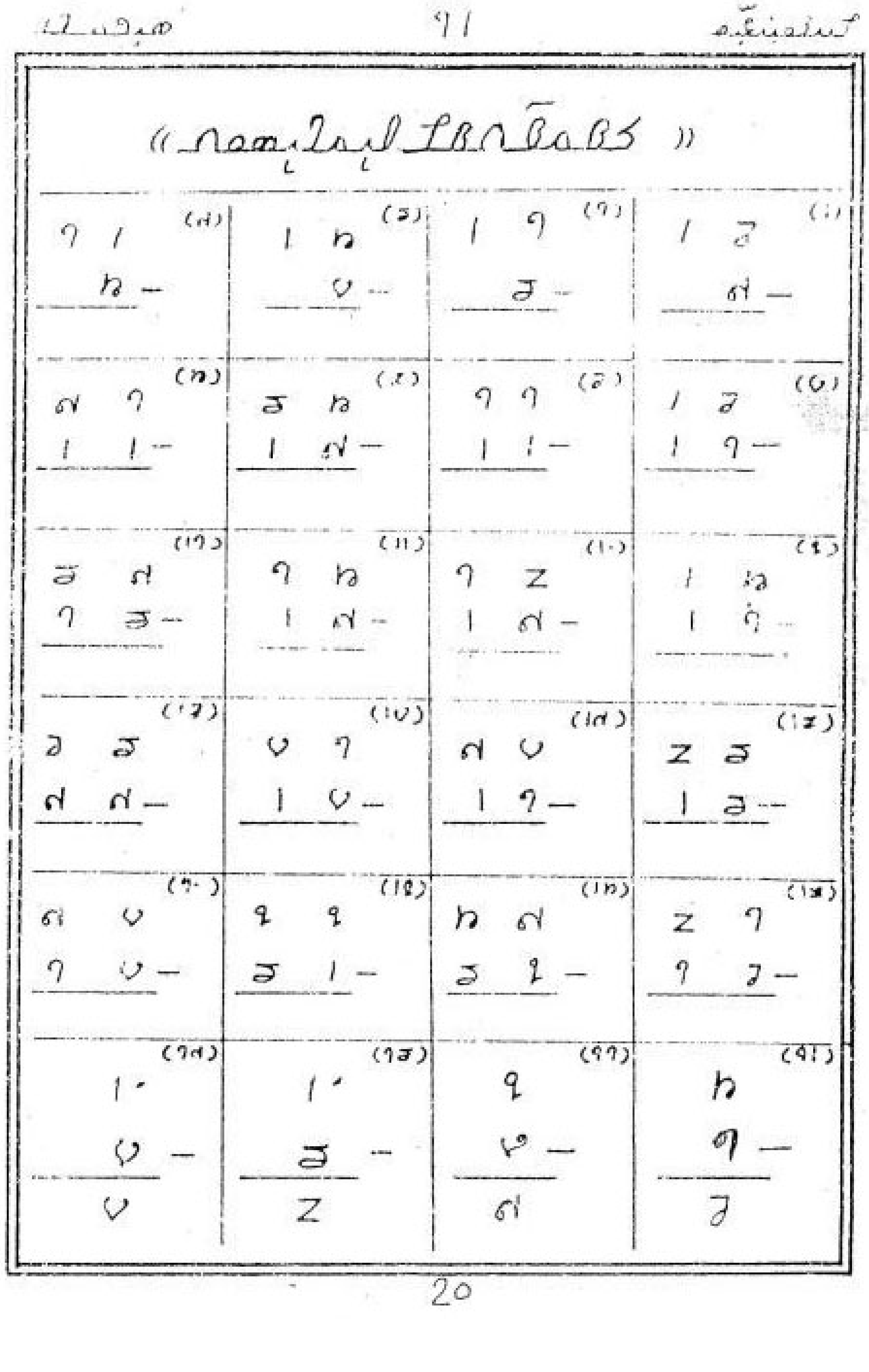


Figure 20: Excerpt from an primary-level arithmetic book written in Rohingya (from Ruwainggya Education Board Myanmar: 21).



Figure 21: The cover page of *Haq-Dar*, a Rohingya language news weekly (December 5, 2002).



Figure 22: The cover page of *Serak*, a Rohingya language news weekly (October 13, 2007).

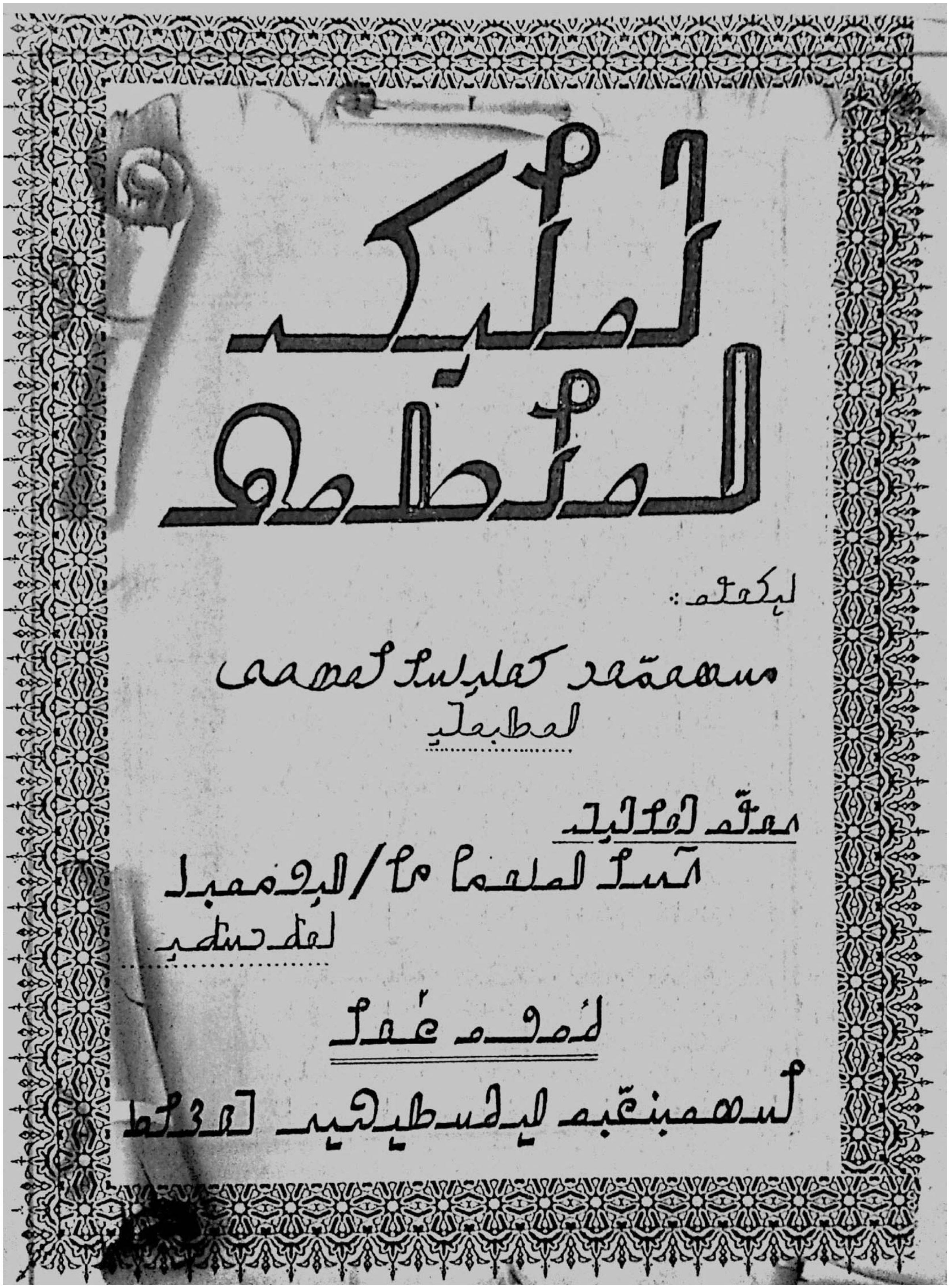


Figure 23: Hand-written cover page of *History of Arakan*. The vertical breaks in the letters 𐴃 TA and 𐴌 RA in the title word 𐴃𐴝𐴌𐴞𐴈 𐴀𐴌𐴝𐴑𐴝𐴕 *tarikh arakan* are stylistic, as are the broken baseline conjuncts through the title.

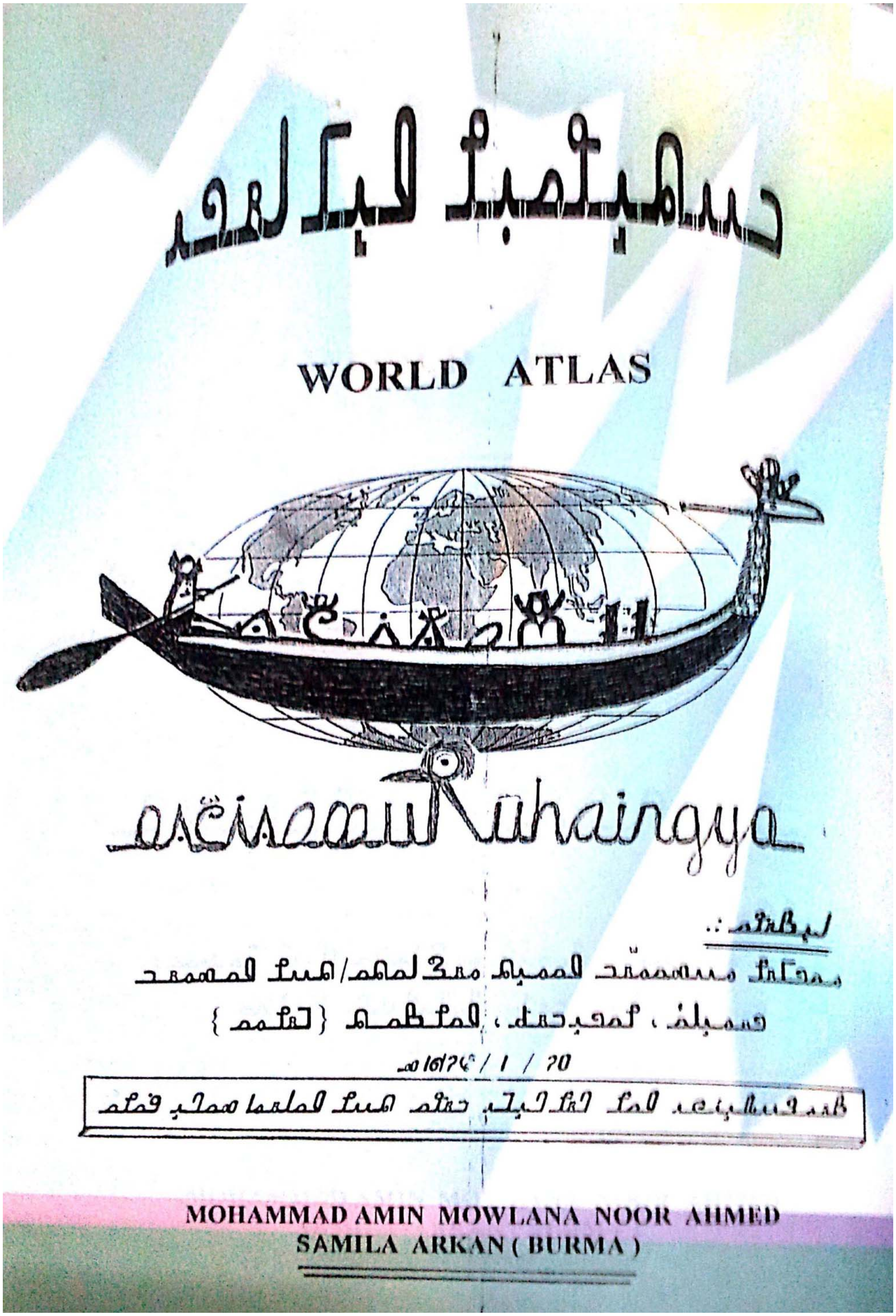


Figure 24: Cover of *World Atlas*.

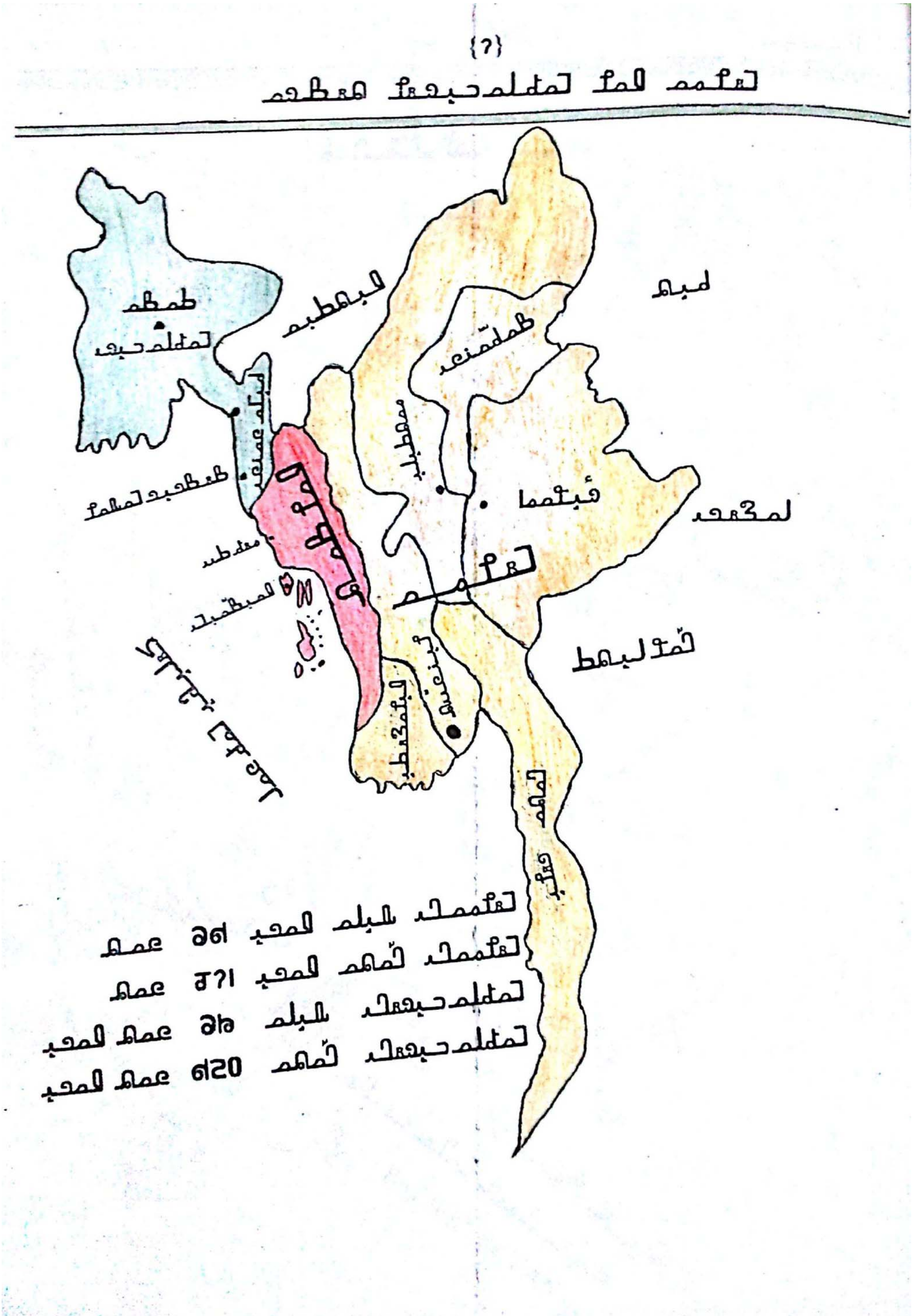


Figure 25: Page from *World Atlas* showing Myanmar (with the Arakan region highlighted) and Bangladesh.



Figure 26: Page from primary school reader. Notice the usage of Arabic punctuation signs.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** ***Proposal to encode Hanifi Rohingya in Unicode***
2. Requester's name: *Anshuman Pandey <pandey@umich.edu>*
3. Requester type (Member body/Liaison/Individual contribution): *Expert contribution*
4. Submission date: *2016-12-31*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *Yes*
   (or) More information will be provided later:

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters): *Yes*
      Proposed name of script: *Hanifi Rohingya*
   b. The proposal is for addition of character(s) to an existing block:
      Name of the existing block:
2. Number of characters in proposal: *50*
3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary	*X*	B.1-Specialized (small collection)		B.2-Specialized (large collection)	
C-Major extinct		D-Attested extinct		E-Minor extinct	
F-Archaic Hieroglyphic or Ideographic				G-Obscure or questionable usage symbols	

4. Is a repertoire including character names provided? *Yes*
   a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document? *Yes*
   b. Are the character shapes attached in a legible form suitable for review? *Yes*
5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
      Anshuman Pandey
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
      Anshuman Pandey
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *Yes*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? *Yes*
7. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? *Yes*

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see Unicode Character Database (http://www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

[1] Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? *No*
   - If YES explain
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? *Yes*
   - If YES, with whom? *Muhammad Noor <mdnoorna@gmail.com>*
     James Lloyd-Williams <james_lloyd-williams@kastanet.org>
     Lorna Priest <lorna_priest@sil.org>
   - If YES, available relevant documents:
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? *Yes*
   - Reference: *See text of proposal*
4. The context of use for the proposed characters (type of use; common or rare) *Common*
   - Reference: *See text of proposal*
5. Are the proposed characters in current use by the user community? *Yes;*
   - If YES, where? Reference: *Speakers of the Rohingya language*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? *N/A*
   - If YES, is a rationale provided?
   - If YES, reference:
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? *Yes*
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? *No*
   - If YES, is a rationale for its inclusion provided?
   - If YES, reference:
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? *No*
   - If YES, is a rationale for its inclusion provided?
   - If YES, reference:
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? *No*
    - If YES, is a rationale for its inclusion provided?
    - If YES, reference:
11. Does the proposal include use of combining characters and/or use of composite sequences? *Yes*
    - If YES, is a rationale for such use provided? *Yes*
    - If YES, reference: *Combining signs for tone marks*
    - Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? *N/A*
    - If YES, reference:
12. Does the proposal contain characters with any special properties such as control function or similar semantics? *No*
    - If YES, describe in detail (include attachment if necessary)
13. Does the proposal contain any Ideographic compatibility characters? *No*
    - If YES, are the equivalent corresponding unified ideographic characters identified?
    - If YES, reference: